UNCLASSIFIED

DTIC FILE COPY

(4)

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

AD-A200 899

DTIC ELECTE NOV 0 4 1988 S D H

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) LIDS-FR-1828	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Lab. for Inf. and Dec. Systems 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Mass. Institute of Technology LIDS/Room 35-410 Cambridge, MA. 02139	7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA. 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-85-K-0329
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

ANALYTIC DEVELOPMENT OF AN EXPERIMENTAL PARADIGM FOR C^3 ORGANIZATIONS

12. PERSONAL AUTHOR(S) Alexander H. Levis, Jeff T. Casey and Anne-Claire Louvet

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 13 May85 TO 12 MAY87	October 1988	140

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The bounded rationality constraint sets an upper limit on the rate with which decisionmakers can process information satisfactorily. This constraints becomes a critical parameter in the design of organizations carrying out command and control functions. Used as a design constraint, it incorporates the notion of avoiding degradation of performance due to excessive workload. An experimental paradigm was developed, a simple computer game for a single decisionmaker, in which subjects were given a limited amount of time to perform a task. Both the amount of time and the task were varied. An information theoretic model of the cognitive workload was used to estimate the workload associated with the tasks. The experimentally determined time threshold at which performance degraded rapidly and the computed cognitive workload led to a value for the bounded rationality constraint for each subject and each task. The distribution of the bounded rationality constraint across subjects for each task was found to be normal. Also, the bounded rationality constraint of each subject as the

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Neil Gerr	22b. TELEPHONE (Include Area Code) (202)696-4321	22c. OFFICE SYMBOL Code 1111SP

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

*U.S. Government Printing Office: 1985—507-047

task changed did not vary significantly. The results of the experimental and analytical investigation may be used in the design of multi-person experiments and in organization design.

OCTOBER 1988 LIDS-FR-1828

ANALYTICAL DEVELOPMENT OF AN EXPERIMENTAL PARADIGM FOR C^3 ORGANIZATIONS

by

Alexander H. Levis
Jeff T. Casey
Anne-Claire Louvet

FINAL REPORT
for the period
13 May 1985 - 12 May 1987

Office of Naval Research
Contract Number N00014-85-K-0329

LABORATORY FOR INFORMATION AND DECISION SYSTEMS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Cambridge, MA 02139

ANALYTICAL DEVELOPMENT OF AN EXPERIMENTAL PARADIGM FOR C^3 ORGANIZATIONS

ABSTRACT

The bounded rationality constraint sets an upper limit on the rate with which decisionmakers can process information satisfactorily. This constraint becomes a critical parameter in the design of organizations carrying out command and control functions. Used as a design constraint, it incorporates the the notion of avoiding degradation of performance due to excessive workload. An experimantal paradigm was developed, a simple computer game for a single decisionmaker, in which subjects were given a limited amount of time to perform a task. Both the amount of time and the task were varied. An information theoretic model of the cognitive workload was used to estimate the workload associated with the tasks. The experimentally determined time threshold at which performance degraded rapidly and the computed cognitive workload led to a value for the bounded rationality constraint for each subject and each task. The distribution of the bounded rationality constraint across subjects for each task was found to be normal. Also, the bounded rationality constraint of each subject as the task changed did not vary significantly. The results of the experimental and analytical investigation may be used in the design of multi-person experiments and in organization design.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

In the period 1982 to 1984, researchers at the Navy Personnel research and Development Center in San Diego, California, carried out a series of experiments on the cognitive demands of command and control decisionmaking (see, for example, Kelley and Greitzer, 1982; Greitzer and Hershman, 1984). Using several versions of simulated Anti-Air Warfare (AAW) operations, they observed a marked performance degradation when the task demands exceeded some limit. Another set of experiments (Greitzer et al., 1984) considered the effect that concurrent tasks had on performance. Results showed that while the two tasks were competing for shared resources, they were not mutually inhibiting. Another observation was that subjects did not use optimal strategies in accomplishing their tasks. The last aspect was explored further by trying to infer strategies and to assess the effect that workload has on the choice of strategies.

The underlying concepts in that experimental effort were very similar, although not identical, to the underlying assumptions of the mathematical theory of organizations that is being developed at the MIT Laboratory for Information and Decision Systems (Levis 1984; 1988). Therefore, a project was undertaken in order to relate the experimental results to the mathematical theory. This report presents the results of the research effort.

The NPRDC work was focused on the single decisionmaker, while the MIT research addressed the organizational problem, i.e., the effect that the cognitive limitations of individual decisionmakers have on organizational performance. However, in order to design an organization and predict its performance, it is necessary that the parameters characterizing the components of the organization be known. In the case of the human decisionmakers, some quantitative expression of the cognitive limitations is needed. The model that has been used is that of the bounded rationality constraint, which states that if the workload rate exceeds some value, rapid degradation of performance occurs. Knowledge of the value or a range of values (with their associated probability distribution) for this threshold could then be used to calibrate the decisionmaker model for used in the algorithms for organizational design and evaluation.

In addressing this issue, the first question is whether such a boundary or threshold exists and whether it is stable across individuals and across tasks. An experimental program was undertaken to investigate this question and place it within the proper framework in cognitive psychology, experimental psychology, and mathematical modeling. The results of the first experiment are the focus of this report.

In Chapter 2, the mathematical model of the Interacting Decisionmaker is presented along with the related models describing the tasks to be performed, the strategies to be used, and a mathematical model for cognitive workload that is based on information theoretic concepts (Levis, 1984). The methodology for evaluating organizational performance -- or the performance of a single decisionmaker -- as a function of the strategies used is also outlined.

The question of workload, as addressed in the behavioral sciences, is discussed in Chapter 3. Specifically, important empirical results and methods from experimental psychology are discussed in the context of determining the bounded rationality constraint of human decisionmakers and in applying the results to command and control processes. This discussion provides the framework for the experimental paradigm described in Chapter 4. The results obtained from carrying out the experiment are presented in Chapter 5. Essentially, the experiment consisted of measuring performance of individual subjects as the amount of time available to carry out the cognitive task was varied.

In Chapters 6 and 7, the results are combined with the mathematical model of information processing and decisionmaking to obtain estimates of the bounded rationality constraint. It is shown, in Chapter 8, that the constraint exists and is stable when minor task changes are made. Furthermore, it is stable across individuals and across tasks.

With these results, which are consistent with the findings at the Navy Personnel Research and Development Center, it is now possible to use the mathematical theory of organizations to design experiments for studying organizational performance in the context of tactical distributed decisionmaking.

2. THE INTERACTING DECISIONMAKER MODEL

The first step in modeling an organizational structure is the modeling of the tasks to be performed by the organization. The second step is to develop an appropriate mathematical model of the organization member. Specifically, this model must incorporate provisions for the variety of interactions that can exist among decisionmakers in an organiztion. These two steps are discussed in this chapter. In addition, the necessary analytical tools are introduced, namely, Petri Nets and N-dimensional information theory. The former is used to describe, rather precisely, the architecture of the decisionmaking model and of the organizations, while the latter is used to model the cognitive workload of the individual decisionmakers.

2.1 PETRI NETS

In this work, only the basic properties of Petri nets are needed to describe the models. In related work for the Office of Naval Research and for the Technical Panel on C^3 of the Joint Directors of Laboratories, several measures of performance (MOPs) of organizations have been obtained using some more advanced concepts from Petri Net theory (Hillion and Levis, 1987; Hillion, 1986). For an introductory treatment of Petri nets as modeling tools, the text by Peterson (1981) is recommended.

Petri Nets are bipartite directed graphs represented by a quadruple (P, T, I, O). By convention, P is the set of one type of nodes, called **places or circle nodes**, and T is the set of the second type of nodes, called **transitions or bar nodes.** Places can depict the presence of signals or represent conditions; transitions can depict processes or events. Consequently, the arcs that connect the nodes that form the graph can only go from one type of node to another - either from a place to transitions, or from a transition to places. The mapping I corresponds to the set of directed arcs from places to transitions, i.e., it defines the input places of the transitions, while the mapping O corresponds to the set of directed arcs from transitions to places; i.e., it defines the output places of each transition. For ordinary Petri Nets - the only type considered here - the mappings I and O take values from the closed set {0,1}; 1 denotes the presence of a link between two nodes, while 0 denotes the absence.

A Petri Net consisting of four transitions and five places is shown in Figure 2.1. Tokens, denoted by dots in places or circle nodes, control the execution of a Petri Net. A **marking** of a Petri Net is a mapping which assigns a non-negative integer number of tokens to each place of the net. Since the number of tokens in a place, in general, is not bounded, there can be an infinite number of markings associated with each net. A Petri Net is said to **execute** when a transitions **fires.** A transition can fire, only if it is **enabled.** For a transition to be enabled, all is input places must contain at least one token each. When a transition fires, it removes one token from each input place and creates a new token in each of the output places of that transition. One can envision a sequence of firings in the Petri Net of Figure 2.1: Let the initial marking consist of a token in the first (leftmost) place. Then the first transition is enabled and it fires. The token in the first place is removed and a token appears in the second place. Now the second transition is enabled: it fires and the token is removed from the second place; a new one appears in the third place, and so on. The execution halts when the fourth transition fires and a token appears on the fifth place.

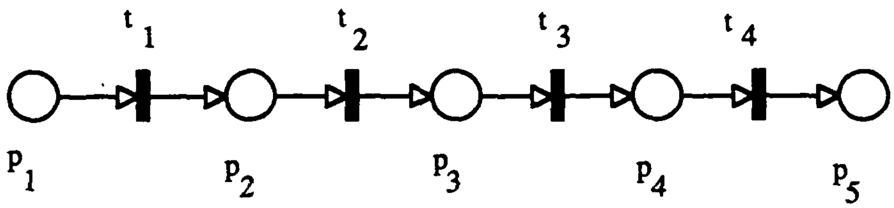


Figure 2.1 A Simple Petri Net

A transition may have more then one output places. When it fires, a token is generated in each output place. However, to model decisionmaking, it is convenient to introduce a special transition, a decision switch, in which the output places represent alternatives. When the decision switch fires, a token is generated in only one of the output places. A decision rule associated with this special transition determines the place in which the token is generated. The rule can be deterministic or stochastic; it can be independent of the attributes of the tokens in the input places or it may depend on them.

A subnet of a Petri Net PN is a Petri Net PN_S with places P_S that are a subset of the places P of the original net and transitions T_S that are a subset of the transitions T of the original net. The input and output mappings, I_S and O_S, are restricted to the arcs between the subsets T_S and P_S. The use of subnets simplifies the graphical representation of complex organizations and allows the depiction of the decisionmaker model at a level of detail appropriate to the problem being solved.

2.2 INFORMATION THEORY

Information theory was first developed as an application in communication theory (Sannon and Weaver, 1949). But, as Khinchin (1957) showed, it is also a valid mathematical theory in its own right, and it is useful for applications in many disciplines, including the modeling of a simple human decisionmaking processes and the analysis of information-processing systems.

There are two quantities of primary interest in information theory. The first of these is **entropy:** given a variable x, which is an element of the alphabet X, and occurs with probability p(x), the entropy of x, H(x), is defined to be

$$H(x) \equiv -\sum_{x} p(x) \log p(x) \tag{2.1}$$

and is measured in bits when the base of the algorithm is two. The other quantity of interest is average mutual information or transmission: given two variables x and y, elements of the alphabets X and Y, and given p(x), p(y), and p(x|y) (the conditional probability of x, given the value of y), the transmission between x and y, T(x:y), is defined to be

$$T(x:y) \equiv H(x) - H_y(x) \tag{2.2}$$

where

$$H_y(x) \equiv -\sum_{y} p(y) \sum_{x} p(x|y) \log p(x|y) \tag{2.3}$$

is the conditional uncertainty in variable x, given full knowledge of the value of variable y.

McGill (1954) generalized this basic two-variable input-output theory to N dimensions by extending Eq. (2.2):

$$T(x_1:x_2: \dots :x_N) \equiv \sum_{i=1}^{N} H(x_i) - H(x_1,x_2,\dots,x_N) \tag{2.4}$$

For the modeling of memory and of sequential inputs which are dependent on each other, the use of the entropy rate, $\overline{H}(x)$, which describes the average entropy of x per unit time, is appropriate:

$$\overline{H}(x) = \lim_{m \to \infty} \frac{1}{m} H[x(t), x(t+1), \ldots, x(t+m-1)] \tag{2.5}$$

The transmission rate, $\overline{T}(x:y)$, is defined exactly like transmission, but using entropy rate in the definition rather than entropy.

The Partition Law of Information (Conant, 1976) is defined for a system with N-1 internal variables, w_1 through w_{N-1}, and an output variable, y, also called w_N. The law states

$$\sum_{i=1}^{N} H(w_i) = T(x:y) + T_y(x:w_1,w_2,\ldots,w_{N-1}) + T(w_1:w_2:\ldots:w_{N-1}:y) + H_x(w_1,w_2,\ldots,w_{n-1},y) \tag{2.6}$$

and is easily derived using information theoretic identities. The left-hand side of Eq. (2.6) refers to the total activity of the system, also designated by G. Each of the quantities on the right-hand side has its own interpretation. The first term, T(x:y), is called **throughput** and is designated G_t. It measures the amount by which the output of the system is related to the input. The second quantity,

$$T_y(x:w_1,w_2,\ldots,w_{N-1}) = T(x:w_1,w_2,\ldots,w_{N-1},y) - T(x:y) \tag{2.7}$$

is called **blockage** and is designated G_b. Blockage may be thought of as the amount of information in the input to the system that is not included in the outpout. The third term, $T(w_1:w_2:\ldots:w_{N-1}:y)$ is called **coordination** and is designated G_c. It is the N-dimensional transmission of the system, i.e., the amount by which all of the internal variables in the system constrain each other. The last term, $H_x(w_1,w_2,\ldots,w_{N-1},y)$, designated by G_n, represents the uncertainty that remains in the system variables when the input is completely known. This **noise** should not be construed to be necessarily undesirable, as it is in communication theory: it may also be thought of as internally-generated information supplied by the system to supplement the input and facilitate the decisionmaking process. The partition law may be abbreviated:

$$G = G_t + G_b + G_c + G_n \tag{2.8}$$

A statement completely analogous to Eq. (2.8) can be made about information rates by substituting entropy rate and transmission rates in Eq. (2.6).

2.3 TASK MODEL

The organization interacts with its environment; it receives signals or messages in various forms that contain information relevant to the organization's tasks. These messages must be identified, analyzed, and transmitted to their appropriate destinations within the organization. From this perspective, the organization, acts as an information user.

Let the organziation receive data from one or more sources (N') external to it. Every τ_n units of time on the average, each source n generates symbols, signals, or messages x_{ni} from its associated alphabet X_n, with probability p_{ni}, i.e.,

$$p_{ni} = p(x_n = x_{ni}) \quad ; \quad x_{ni} \in X_n \quad i = 1, 2, \ldots, \gamma_n \tag{2.9}$$

$$\sum_{i=1}^{\gamma_n} p_{ni} = 1 \quad ; \quad n = 1, 2, \ldots, N' \tag{2.10}$$

where γ_n is the dimension of x_n. Therefore, $1/\tau_n$ is the mean frequency of symbol generation from source n.

The organization's task is defined as the processing of the input symbols x_n to produce output symbols. This definition implies that the organization designer knows a priori the set of desired responses Y and, furthermore, has a function or table $L(x_n)$ that associates a desired response or a set of desired responses, elements of Y, to each input $x_{ni} \in X_n$.

It is assumed that a specific complex task that must be performed can be modeled by N' sources of data. Rather than considering these sources separately, one supersource, composed of these N' sources, is created. The input symbol x' may be represented by an N'-dimensional vector with each source corresponding to a component of this vector, i.e.,

$$\underline{x}' \equiv (x_1, x_2, \ldots, x_{N'}) \quad ; \quad \underline{x}' \in X \tag{2.11}$$

To determine the probability that symbol $\underline{x}'_j$ is generated, the independence between components must be considered. If all components are mutually independent, then p_j is the product of the probabilities that each component of $\underline{x}'_j$ takes on its respective value from its associated alphabet., Eq. (2.12). If two or more components are probabilistically dependent on each other,

$$p_j = \prod_{n=1}^{N'} p_{nj} \tag{2.12}$$

but as a group are mutually independent from all other components of the input vector, then these dependent components can be treated as one supercomponent, with a new alphabet. Then a new input vector, $\underline{x}$, is defined, composed of the mutually independent components and these super-components.

This model of the sources implies synchronization between the generation of the individual source elements so that they may, in fact, be treated as one input symbol. Specifically, it is assumed that the mean interarrival time τ_n for each component is equal to τ. It is also assumed that the generation of a particular input vector, $\underline{x}_j$, is independent of the symbols generated prior to or after it.

The last assumption can be weakned, if the source is a discrete stationary ergodic one with constant interarrival time τ that could be approximately by a Markov source. Then the information theoretic framework can be retained (Hall and Levis, 1983).

The vector output of the source is partitioned into groups of components that are assigned to different organization members. The j-th partition is denoted by $\underline{x}^j$ and is derived from the corresponding partition matrix π^j which has dimension n_j x N and rank n_j, i.e.,

$$\underline{x}^j = \pi^j \, \underline{x}. \tag{2.13}$$

Each column of π^j has at most one non-zero element. The resulting vectors x^j may have some, all, or no components in common.

The set of partitioning matrices $\{\pi_1, \pi_2, ..., p_n\}$ shown in Figure 2.2 specify the components of the input vector received by each member of the subset of decisionmakers that interact directly with the organization's environment. These assigmnets can be time invariant or time varying. In the latter case, the partition matrix can be expressed as

$$\pi^j(t) = \begin{cases} \pi^j_o & \text{for } t \in \{T\} \\ 0 & \text{for } t \notin \{T\} \end{cases} \tag{2.14}$$

The times {T} at which a decisionmaker receives inputs for processing can be obtained either through a deterministic (e.g., periodic) or a stochastic rule. The question of how to select the set of partition matrices, i.e., design the information structure between the environment and the organization, has been addressed by Stabile and Levis (1984); Stabile (1981).

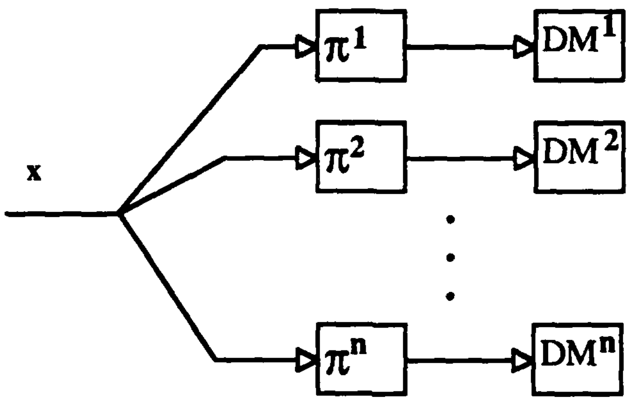


Figure 2.2 Information Structures for Organizations

2.4 THE DECISIONMAKER MODEL

The basic model of the memoryless decisionmaker with bounded rationality is based on the hypothesis of F.C. Donders (1983) that information processing is done in stages. Specifically, it is assumed that the two stages are (a) situation assessment (SA), and (b) response selection (RS), which correspond to March and Simon's (1958) two stage process of discovery and selection. The structure of this model, shown in Figure 2.3, has been extended to include interactions with other organization members, as well as memory. The extended model is shown in Figure 2.4.

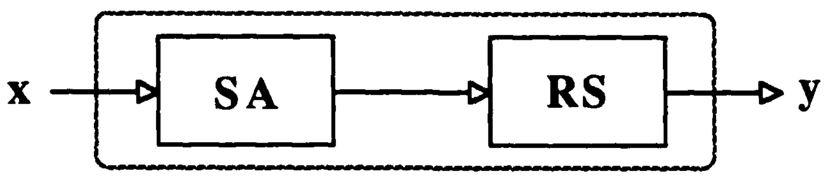


Figure 2.3 Two-Stage Model

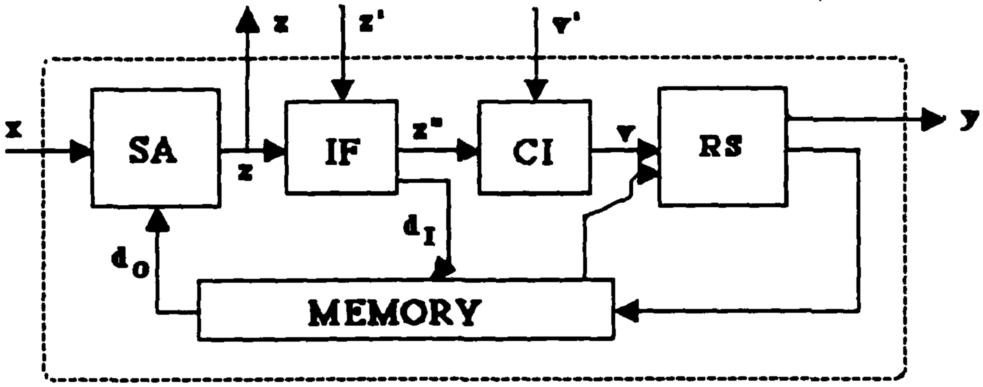


Figure 2.4 The Interacting Decisionmaker with Memory

The DM receives signals $x \varepsilon X$ from the environment with interarrival time τ. A string of signals may be stored first in a buffer so that they can be processed together in the situation assessment (SA) stage. The SA stage contains algorithms that process the incoming signals to obtain the assessed situation z . The SA stage may access the memory or internal data base to obtain a set of values d_O. The assessed situation z may be shared with other organization members; currently, the DM may receive the supplementary situation assessment z' from other parts of the organization; the two sets z and z' are combined in the information fusion (IF) processing stage to obtain z". Some of the data (d_I) from the IF process may be stored in memory.

The possibility of receiving commands from other organization members is modeled by the variable v'. A command interpretation (CI) stage of processing is necessary to combine the siatuion assessment z" and v' to arrive at the choice v of the appropriate strategy to use in the response selection (RS) stage. The RS stage contains algorithms that produce outputs y in response to the situation assessment z" and the command inputs. The RS stage may access data from, or store data in memory (Hall and Levis, 1983).

In this report, only the memoryless case is considered. Consequently, the general model reduces to the one shown in Figure 2.5, where the Petri Net formalism has been used.

A more detailed description of the model is obtained, if the internal structure of the SA and RS stages is considered. The situation assessment stage consists of a set of U algorithms (deterministic or not) that are capable of producing some situation assessment z". The choice of algorithms is achieved through specification of the internal variable u in accordance with the situation assessment strategy p(u), or p(u|x), if a decision aid (e.g., a preprocessor) is present. A second internal decision is the selection of the algorithm in the RS stage according to the response selection strategy p(v|z,v'). The two strategies, when taken together, constitute the internal decision strategy of the decisionmaker.

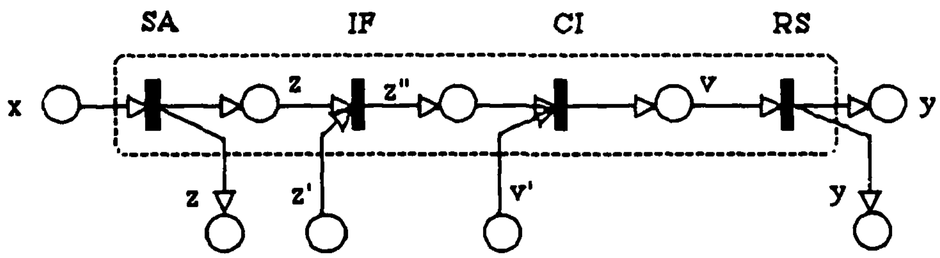


Figure 2.5 The Memoryless Interacting Decisionmaker Model

The subnets representing the SA and the RS stages are shown in Figure 2.6. Note the presence of decision switches in place of the regular transitions to indicate that only one of the output places can receive a token at each firing.

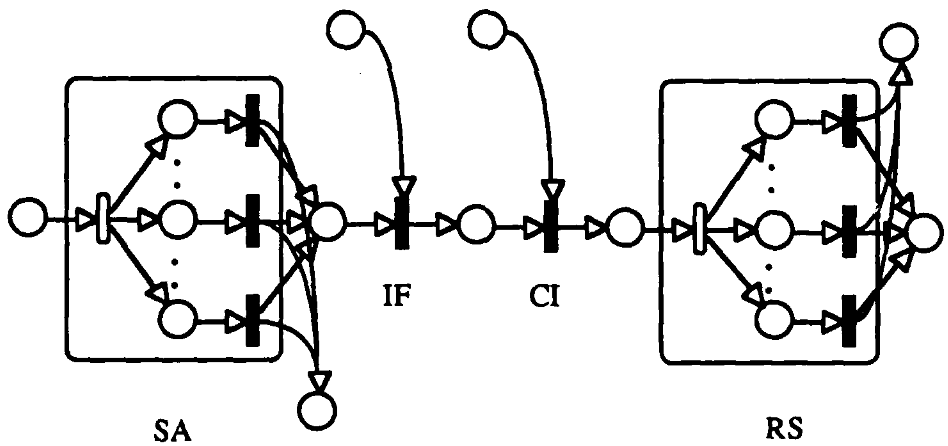


Figure 2.6 The SA and the RS Subnets

2.5 WORKLOAD

The analytical framework presented in Section 2.2, when applied to the single interacting decisionmaker with deterministic algorithms in the SA and RS stages, yields the four aggregate quantities that characterize the information processing and decisionmaking activity within the DM:

Throughput:

$$G_t = T(x,z',v':z,y) \tag{2.15}$$

Blockage:

$$G_b = H(x,z',v') - G_t \tag{2.16}$$

Internally generated information:

$$G_n = H(u) - H_z(v) \tag{2.17}$$

Coordination:

$$G_c = \sum_{i=1}^{U} [p_i g_c^i(p(x)) + \alpha_i H(p_i)] + H(z) + g_c^{IF}(p(z,z')) + g_c^{CI}(p(z,v')$$

$$+ \sum_{j=1}^{V} [p_j g_c^i (p(z|v=j)) + \alpha_j H(p_j] + H(y) + H(z) + H(z) + Hz,v)$$

$$+ T_z(x':x') + T_z(x',z':v') \tag{2.18}$$

The expression for G_n shows that it depends on the two internal strategies p(u) and p(v|z) even though a command input may exist. This implies that the command input v' modifies the DM's internal decision after p(v|z) has been determined.

In the expressions defining the system coordination, p_i is the probability that algorithm f_i has been selected for processing the input x and p_j is the probability that algorithm h_j has been selected, i.e., u=i and v=j. The quantities g_c represent the internal corodinations of the corresponding algorithms and depend on the probability distribution of their respective inputs; the quantities α_i, α_j are the number of internal variables of the algorithms f_i and h_j, respectively. Finally, the quantity **H** is the entropy of a binary random variable that takes one of its two values with probability p.

$$H(p) = - p \log_2 p - (1-p)\log_2(1-p) \tag{2.19}$$

Equations (2.15) to (2.18) determine the total activity G of the decisionmaker according to the partition law of information, Eq. (2.6). The activity G can be evaluated alternatively as the sum of the marginal uncertainties of each system variable. For any given internal decision stragety, G and its component parts can be computed.

Since the quantity G may be interpreted as the total information processing activity of the system, it can serve as a *surrogate for the workload* of the organization member in carrying out his decisionmaking task.

The qualitative notion that the rationality of a human decisionmaker is not perfect, but is bounded (March, 1978), has been modeled as a constraint on the total activity G. The specific form for the constraint has been suggested by the empirical relation

$$t = c_1 + c_2 G_t$$

where t is the average reaction time, i.e., the time between the arrival of the input and the generation of an output y. It is assumed that the decisionmaker must process his inputs at a rate that is at least equal to the rate with which inputs arrive. The latter has been model by τ, the mean symbol interarrival time:

$$t = c_1 + c_2 G_t \leq \tau$$

or

$$\frac{1}{c_2} t = \frac{c_1}{c_2} + G_t \leq \frac{1}{c_2} \tau$$

The modeling assumptions in this work are that

$$\frac{c_1}{c_2} = G_b + G_n + G_c$$

and that c_2 does not depend on p(x). Then, the bounded rationality constraint takes the form

$$G = G_t + G_b + G_n + G_c \leq \frac{1}{c_2} \tau = F \tau \tag{2.20}$$

where F can be considered as a rate of total activity and is measured in bits per second. Inequality (2.20) represents a mathematical expression of only one aspect of bounded rationality. Many other formulations are possible.

Weakening the assumption that the algorithms are deterministic changes the numerical values of G_n and of the coordination term G_c (Chyen and Levis, 1985). If memory is present in the model, then additional terms appear in the expressions for the coordination rate and for the internally generated information rate (Hall and Levis, 1983).

2.6 MEASURES OF PERFORMANCE

As stated in Section 2.3, it is assumed that the designer knows a priori the set of desired responses Y to the input set X. One measure of performance (MOP) of the organization that reflects the degree to which the actual response matches the desired response can be computed as shown in Figure 2.7.

The decisionmaker's actual response y can be compared to the desired response y' and a cost is assigned using the cost function d(y,y'). If this function is a binary one, i.e.,

$$d(y,y') = \begin{cases} 0 & \text{if } y = y' \\ 1 & \text{if } y \neq y' \end{cases} \tag{2.21}$$

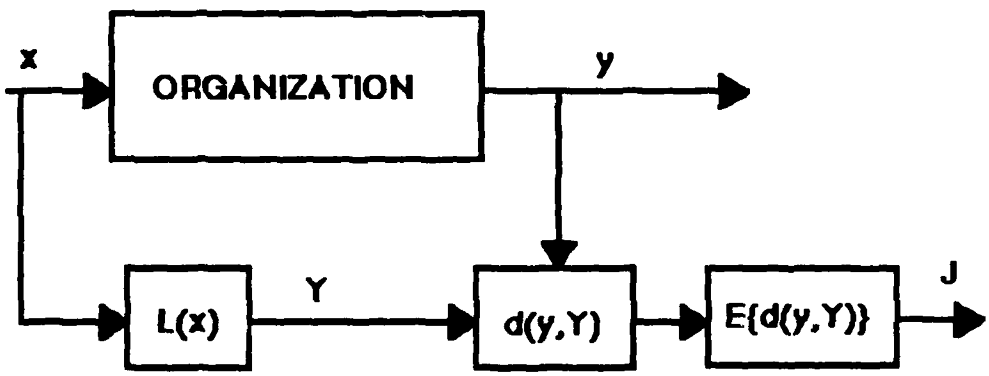


Figure 2.7 Performance Evaluation of an Organization

then the expected value of this cost denotes the probability that the wrong decision is made, i.e., it is the probability of error.

In general, however, there is a cost c_{ij} associated with selection $y_i \varepsilon Y$ when the desired response is $y'_j \varepsilon Y'$:

$$C_{ij} = d(y_i, y'_j) \tag{2.22}$$

so that

$$J = \sum_j p(x_j) \sum_i c_{ij}\, p(y_i|x_j) \tag{2.23}$$

where y'_j is the desired response to task x_j. This measure of performance can be interpreted as a measure of the **accuracy** of the response, to the extent that a cost is associated with the degree with which the actual decision deviates from the desired one.

This class of performance measures, described generically by (2.23), is not the only one that has been considered. In related work (Andreadakis and Levis, 1987), measures of performance that address time have been modeled and analyzed.

2.7 PERFORMANCE-WORKLOAD LOCUS

A useful way for describing the properties of the decisionmaker model, which is generalizable to the properties of an organization, is through the performance workload locus. In the case of a single performance measure, the accuracy measure J, and a single decisionmaker with workload G, a two dimensional space is defined with ordinate J and abscissa G. The locus is constructed by considering the functional dependence of J and G on the internal decision strategies of the single decisionmaker.

Let an internal strategy for a given decisionmaker be defined as pure, if both the situation assessment strategy p(u) and the response selection strategy p(v|z) are pure, i.e., an algorithm f_i is selected with probability one and an algorithm h_j is selected also with probability one when the situation is assessed as being z_k:

$$D_k = \{p(u{=}i) = 1 \;\; ; \;\; \{p(v{=}j|z{=}z_k) = 1\}\} \tag{2.24}$$

for some i, some j, and for each z_k element of the alphabet Z. There are n possible pure internal strategies,

$$n = U \cdot V^M \tag{2.25}$$

where U is the number of f algorithms in the SA stage, V the number of h algorithm in the RS stage and M the dimension of the set Z. All other internal strategies are mixed (Boettcher and Levis, 1982) and are obtained as convex combinations of pure strategies:

$$D(p_k) = \sum_{k=1}^{n} p_k D_k \tag{2.26}$$

where the weighting coefficients are probabilities.

Corresponding to each $D(p_k)$ is a point in the simplex

$$\sum_{k=1}^{n} p_k = 1, \quad p_k \geq 0 \ \forall\, k \tag{2.27}$$

The possible strategies for an individual DM are elements of a closed convex polyhedron of dimension n-1 whose vertices are the unit vectors corresponding to pure strategies.

The total activity G, the surrogate for the cognitive workload, is a convex function of the decision strategy, i.e.,

$$G\,(D(p_k)) \geq \sum_{k=1}^{n} p_k\, G_k \tag{2.28}$$

where G_k is the workload that results when the pure strategy D_k, given by Eq. (2.24), is used.

The accuracy measure J can be related to the decision strategies in a similar manner. Corresponding to each pure strategy D_k is a value of the performance measure, denoted by J_k. Since each strategy is a convex combination of pure strategies, the value of J for an arbitrary $D(p_k)$ is given as a convex combination of the values of J_k, i.e.,

$$J(D(p_k) = \sum_{k=1} p_k J_k \tag{2.29}$$

The two expressions (2.28) and (2.29) can be used now to characterize the locus of points in the (J,G) space that describe the decisionmaker.

<u>Example</u>: Consider first the case of two pure strategies, D_1 and D_2. This would correspond to the case where the decisionmaker can choose only between two different algorithms f in the SA stage, as shown in Figure 2.8. The strategy space for this case can be parameterized as follows: Any strategy, D, can be expressed as

$$D = p_1 D_1 + p_2 D_2 \tag{2.30}$$

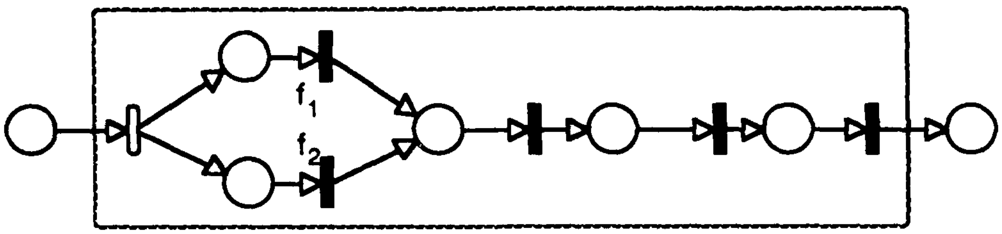


Figure 2.8 The Petri Net for the Example

where

$$p_1 + p_2 = 1$$

in accordance with (2.26) and (2.27). Let

$$p_1 = 1 - \delta \quad \text{and} \quad p_2 = \delta$$

and let

$$0 \leq \delta \leq 1.$$

Then, (2.30) can be rewritten as

$$D = (1\text{-}\delta)\, D_1 + \delta D_2 \tag{2.31}$$

The strategy space can be described by the parameter δ: it is the line segment [0,1], as shown in Figure 2.9, with the point 0 corresponding to pure strategy D_1, point 1 to pure strategy D_2, and all points in between to all the mixed strategies.

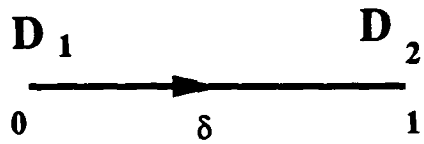


Figure 2.9 Strategy Space for Example

Then, it follows from (2.28) and (2.29) that

$$G(D(p_k)) = G(D(\delta)) \geq (1\text{-}\delta)G_1 + \delta\, G_2 \tag{2.32}$$

and

$$J(D(p_k)) = J(D(\delta)) = (1\text{-}\delta)\, J_1 + \delta\, J_2 \tag{2.33}$$

Equations (2.32) and (2.33) are parametric in δ and result in the locus shown in Figure 2.10. The relative position of the end points (J_1,G_1) and (J_2,G_2) is problem specific; it is not true that smaller workload leads to worse performance, as Figure 2.10 indicates.

In the general case, there are n pure strategies, as given by Eq. (2.25). Then, the P-W locus is constructed as follows:

First, the values of (J_i,G_i) for the n pure strategies are determined. This corresponds to evaluating the performance and the workload for the values of p_k, Eq. (2.27), that correspond to the vertices of the strategy space. The result is a set of n points in the two-dimensional P-W space.

Then, the binary variations between each possible pair of pure strategies are considered. This corresponds to the mapping of the edges of the strategy space. For example, consider pure strategies D_i and D_k: then

$$D = (1\text{-}\delta)\, D_i + \delta D_k$$

for all combinations (i,k) where i=1,...,n and k=1,...,n and for which $i \neq k$. By varying δ from 0 to 1, the loci ($J_{ik}(\delta)$, $G_{ik}(\delta)$) are obtained. These are convex lines joining the two boundary points, as shown in Figure 2.10. These binary loci are quite useful, since they define the minimum workload locus for any feasible value of J.

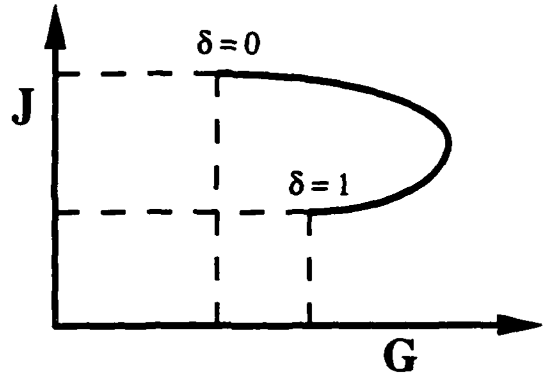


Figure 2.10 Performance-Workload Locus for Example

The third step consists of considering, successively, the binary variation between all possible binary strategies until all mixed strategies are accounted for. The result is a locus such as the one shown in Figure 2.11 for the case when there are three pure strategies. The corresponding strategy space, for this case, is shown in Figure 2.12.

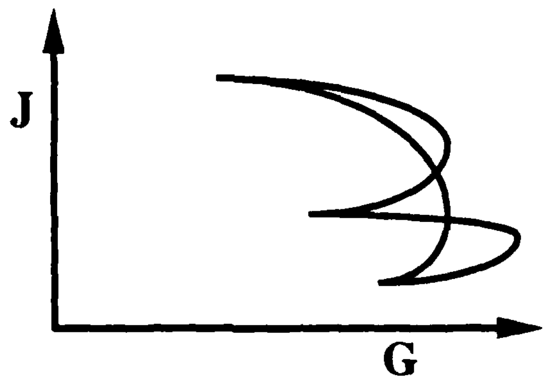


Figure 2.11 Performance-Workload Locus for the Case of Three Pure Strategies

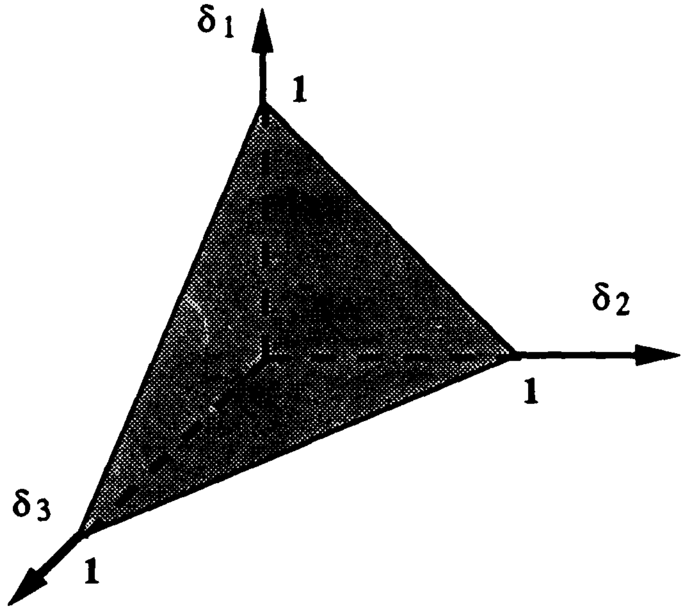


Figure 2.12 Strategy Space for the Case of Three Pure Strategies

Thus, the decisionmaker model can be considered as a system that maps the strategy locus, the simplex defined by Eq. (2.27), into the Performance-Workload (J,G) locus. Any change in the algorithms f or h, or the functions in IF and CI, or the input x will affect the mapping.

In the next chapter, a review of relevant material from experimental psychology is presented in order to set the stage for the description and analysis of the experimental paradigm.

3. WORKLOAD AND BEHAVIORAL DECISION THEORY

This chapter presents selected findings and methods from experimental psychology which are relevant to the analysis and evaluation of C^2 organizations; they provide a basis for the experimental paradigm presented in this report. The issues raised apply specifically to:

(1) modeling the information processing algorithms used by individual decisionmakers;
(2) evaluating the cognitive workloads associated with these algorithms via the information theoretic surrogate for workload proposed by Boettcher and Levis (1982); and
(3) testing, extending, and applying the workload surrogate.

These issues come primarily from three areas. *Behavioral decision theory* is discussed in terms of its implications for modeling situation assessment and response selection algorithms. *Cognitive psychology*, specifically models of attention, is discussed as it relates to the theoretical underpinnings of the concept of workload. Finally, literature from *human performance* is reviewed in order to assess the state of the art in measurement of workload. Within each of these areas a few key references are suggested for further reading.

3.1 POTENTIAL CONTRIBUTIONS FROM BEHAVIORAL DECISION THEORY

The behavioral decision literature can be partitioned roughly into three areas: judgment and decision under certainty, intuitive statistics/heuristics and biases, and decision under risk. These three areas are described briefly to provide context for the ensuing discussion. Two of the dominant theories from this literature are suggested as possible means for identifying a broadly applicable set of possible algorithms for modeling the situation assessment process and response selection processes. Then, experimental research is reviewed which bears on three issues relevant to modeling decision behavior in the C^3 context. The first issue concerns whether strategy selection can be predicted from knowledge of the decision maker's risk attitude. The remaining issues concern two highly salient features of tactical battle management environments: time pressure and dynamic evolution of scenarios.

3.1.1 Overview of the field

Behavioral decision theory is the study of how people make judgments about the world and their own preferences, and how these judgments are combined and compared to make decisions. The field grew originally out of economics. Its origin is usually traced to von Neumann and Morgenstern's (1947) axiomatization of expected utility theory. However, cognitive (and social) psychology has come to play an increasing crucial role in recent years. Each of the field's three major areas will be discussed in turn and related to the modeling of the individual decision maker.

Judgment and decision under certainty is the study of how people combine multiple sources of information into a judgment along a single dimension. This area has been driven by the *information integration theory* (not to be confused with information theory) of Anderson (1974; 1983). Although this theory has been used most often to describe judgment under certainty, it is equally applicable to many judgment tasks involving uncertainty (cf. Anderson and Shanteau, 1970). Information integration theory is discussed below in more detail in the context of mathematical models of situation assessment.

Intuitive statistics/heuristics and biases deals with simplifying strategies people use for estimating and revising probabilities and making predictions and inferences. Investigations typically take the form of comparison of individuals' judgments with normative rules, particularly probability theory. The standard conclusion is that, because of bounded rationality and ignorance of normative theory, people use simplifying judgmental heuristics that in some cases lead to large and systematic errors. Although there is little doubt that these errors are real, the practical difficulty with this research is that the heuristics that are identified (e.g., availability and representativeness) are little more than restatements of the phenomena they are intended to explain. For example, the *availability* heuristic involves judging the probability of an event according to the ease with which instances where the event did occur can be remembered. This heuristic is incomplete; it cannot be substituted for a normative probability model without additional assumptions. Nonetheless, where situation assessment tasks involving processing or generating probability estimates are concerned, this research suggests a number of starting points for algorithmic modeling. See Kahneman, Slovic and Tversky (1982) for a detailed "catalog" of these heuristics and biases.

Decision under risk is the oldest branch of behavioral decision theory and the branch most heavily influenced by economics. The seminal work of von Neumann and Morgenstern dealt with this branch. Until recently this branch was dominated totally by expected utility theory and its more complex variants (e.g., Karkarmar's, 1978, subjective expected utility theory and Kahneman and Tversky's, 1979, prospect theory). Although this lineage of theory has the virtue of mathematical elegance, numerous empirical demonstrations have cast doubt on the validity of the underlying behavioral assumptions (see Schoemaker, 1982, for a review). Due to the influence of cognitive psychology, attempts to model risky decision behavior using algorithms and empirically based behavioral assumptions have begun to emerge (e.g., Lopes, in press; Payne, 1982). Such models are called *process* or *procedural* models. and, although they typically deal with greatly simplified tasks, they are easily amenable to modeling via the workload surrogate.

The ideas to be discussed in this section are by no means all that behavioral decision theory has to offer the C^3 organizational design researcher. Excellent reviews of behavioral decision theory are provided by Slovic, Fischhoff, and Lichtenstein (1977), and Einhorn and Hogarth (1981). An excellent and authoritative text is Hogarth (1980). Hogarth's text is narrow in that it focuses primarily on the cognitive processes of individual decision makers and how pitfalls (cognitive biases) can be avoided. Wright (1985) is a collection of primarily original papers which offers a much broader descriptive perspective on decision behavior.

3.1.2 General models of situation assessment and response selection

The theories to be discussed in this section potentially provide general, behaviorally valid models of situation assessment and response selection processes. These models are most appropriate for cases in which little context- or task-specific information is available as to what sorts of algorithms decision makers may actually employ.

Situation Assessment. *Information integration theory* is a family of simple algebraic models of how information on various dimensions is combined to produce a judgment on a single dimension (Anderson, 1974). The input dimensions can be defined qualitatively (ordinarily) or quantitatively. In a C^3 situation assessment context involving aircraft identification, input dimensions might be speed, direction, and whether radio contact has been made. The resulting subjective assessment might be likelihood that the aircraft is hostile.

Variations of an *averaging* model (i.e., average or weighted average) or an even simpler *adding* model (i.e., sum or weighted sum) are descriptively accurate in many situations. It is taken for granted that people do not perform conscious mental arithmetic in order to produce averaging judgments. However, a plausible algorithmic model has been proposed that explains how people might produce judgments consistent with the models (Lopes, 1982). Lopes model is, to a substantial degree, a model of the heuristics people use to do certain kinds of mental arithmetic.

Response Selection. Kahneman and Tversky's (1979) prospect theory, despite its numerous limitations (cf. Schneider and Lopes, 1987), is the most comprehensive and generally applicable theory of choice behavior (response selection) yet proposed. It is primarily a mathematical theory in the spirit of expected utility theory. It concerns how probability and value information is evaluated and combined in order to allow direct comparison between alternative courses of action or responses. Some of its shortcomings in terms of modeling C^3 response selection are:

(1) alternatives must be defined on two dimensions only (probability and value) and must, technically, have only two possible outcomes each;

(2) Kahneman and Tversky do not provide a specific equation for their value function which maps objective values onto psychological values; and

(3) It is not a process or procedural theory; as with information integration theory, it is assumed that people do not actually execute the computations the theory suggests.

However, in the case of choice behavior, it is even less clear that the underlying algorithms resemble the mechanics of the theory in any way.

3.1.3 Individual differences in risk taking

Everyday experience suggests that propensity to take risks is one of the things that gives people their individuality. Some people seem to seek out risks at every turn, while others try equally hard to avoid them. If a straightforward method (e.g., a paper and pencil questionnaire) were available for classifying individuals with respect to risk attitude, this information could be used to predict what sorts of algorithms a given decision maker is more likely to use in a given task. For example, given a set of strategies that are equally effective *on average,* a risk seeker might preferentially use a "quick and dirty" strategy that offers a chance of an unusually successful outcome. A risk avoider, on the other hand, might select a safe strategy that minimizes the chances of failure.

Despite its tremendous common sense appeal, validating risk attitude as a personality trait has proved to be quite difficult. Early attempts to develop valid and reliable questionnaire measures failed (Slovic, 1962). For example, the Choice Dilemma Questionnaire (CDQ) has been a popular means of measuring risk taking propensity. It was the measure used in the discovery of well-known "risky shift" phenomenon of group decision making (Cartwright, 1973) and in many follow-up studies. However, Bazerman (in press) has shown that the CDQ is likely to show large biases because of the way the questions are framed. It appears that an entire body of literature was predicated on an invalid risk measure.

A recurring theme in psychological research is that attempts to identify stable traits which bear out common sense notions about individual differences often lead instead to an awareness of the tremendous flexibility and adaptivity of behavior. Behavior is determined more by environmental demands and less by personal traits than people commonly believe. This discrepancy has been labeled the "fundamental attribution error" (Nisbett and Ross, 1980).

Lopes (in press) has proposed a three part model of risk taking which reflects adaptation to the situation at hand, while also allowing identification of stable individual differences in risk taking. According the Lopes model, while nearly everyone is concerned with achieving outcomes that are at least *acceptable* in risky decisions (i.e., meeting their *aspiration level),* people differ in the weight they place on *security* and *potential.* Aspiration level is situation-specific; it depends upon what the available alternatives have to offer. Security refers to the probability that a risky alternative will allow the decision maker to avoid a ruinous or very damaging loss. Potential refers to the probability of "winning big" -- that is, achieving an outcome much higher than the aspiration level. Most people give more weight to security than potential. However, a substantial proportion focus on potential, and make correspondingly very risky choices.

In order to measure risk taking propensity, the Lopes method uses multi-outcome probability distributions involving money that differ in terms of security and potential, as well as other factors. Although this method has a firm foundation in psychological theory, it has *not* been validated formally as a psychometric test. An additional caveat is that Lopes and Casey (1987) found situational effects in risk taking that were presumably due to (intra-individual) *shifts* in attention to security and potential. Thus an individual's "trait" tendency to take or avoid risk is modulated not only by aspiration level but by other situational ("state") factors, as well.

3.1.4 Decision making under time pressure

Time pressure is one of the most salient features of decision making in the context of tactical battle management. Yet this feature has been all but ignored in the behavioral decision literature. The reason may have to do with decision researchers' preoccupation with identifying biases in judgment. This preoccupation is discussed by Christensen-Szalanski and Beach (1984). Showing that a bias occurs under time pressure is less conclusive evidence for the importance and generality of the bias than demonstrating that it occurs even when processing time is unlimited.

Although time pressure is used frequently in investigations of basic cognitive processes (attention, memory, etc.) this work is even less relevant, because of the highly simplified and artificial nature of the tasks (e.g., reliance on nonsense syllables as stimuli). The limitations created by this artificiality are discussed by Neisser (1976). Further, the time pressure is not in the form of a temporal "window of opportunity." Rather, subjects are given plenty of time, but instructed to respond as quickly as possible without sacrificing accuracy.

A few studies of time pressure do exist in the behavioral decision literature. These studies have clear implications for modeling of the situation assessment and response selection processes. The general conclusion of these studies is that, under time pressure, people process only a portion of the information they would process normally. Further, they filter the information, so that the information that is processed is more important than that which is not processed (Ben Zur and Breznitz, 1981; Wright, 1974; Wright and Weitz, 1977).

Hogarth (1975) proposed a general mathematical model for predicting how decision makers select decision strategies when faced with multi-attribute alternatives. According to this model, the decision maker selects the strategy which offers the optimal trade off between the cost of decision time and the cost of errors. This model attempts to explain the counter-intuitive empirical finding that less difficult decisions do not always take less time. It appears that this model could be modified to reflect the impact of a temporal window of opportunity on the cost of decision time. A model of this type might be useful in defining the probability distribution over possible strategies as a function of the degree of time pressure.

The most sophisticated work to date in the area of decision behavior under time pressure was conducted by Payne and colleagues (Bettman, Johnson, and Payne, 1986; Payne, Bettman, and Johnson, 1986). This research has a number of points of tangency with C^3 research.

First, these researchers proposed and tested experimentally an *objective* measure of cognitive effort. This measure bears a closer resemblance to the workload surrogate of Boettcher and Levis (1982) than any other popular method. Effort is estimated as the sum or weighted sum of the *elementary information processes* involved in executing an algorithm. This approach is in the tradition of Newell and Simon (1972). Elementary information processes are READS, ADDITIONS, COMPARISONS, PRODUCTS, DIFFERENCES, and ELIMINATIONS. This approach was shown to predict response time more accurately than several alternative methods. The authors conclude that "...these results imply that a small number of simple operators can be viewed as the fundamental components from which decision rules are constructed" (Bettman et al., 1986 p. 35).

In testing the predictions of the elementary information processes method, Payne and colleagues faced a problem similar to that faced in testing the workload surrogate (see Louvet, Casey, and Levis; 1988). That is, it was necessary to constrain the decision strategies used by experimental subjects. Toward this end, they employed an IBM PC-based information acquisition system which is generally available. This system, called "Mouselab", presents information displays from which subjects can access individual pieces of information using a mouse. The information acquisition process can be controlled or monitored precisely, in order to reduce the range of possible strategies or infer what strategies are being used. For example, Bettman et al. trained subjects to use six different strategies. For the actual experiment, subjects were told for each set of trials which strategy to use. The software permitted subjects to use only this strategy; errors in adherence to the strategy were signaled to the subject and recorded for later analysis.

The Payne et al. study is quite unique in that shifts in strategy as a function of time pressure and other task variables were monitored in real time. In this study, subjects were permitted to use whatever strategy or strategies they desired. Some of the major results were that, under moderate time pressure, subjects processed less information and processed it at a higher rate. Under severe time pressure, qualitative shifts in decision strategies also occurred. Adaptivity to task characteristics increased with experience. Payne et al. concluded that people have a repertoire of heuristic strategies available to them and that, using knowledge of the task structure and the degree of time pressure, they are able to choose heuristics which are acceptable in terms of effort (workload and timeliness) and accuracy.

3.1.5 Dynamic evolution of decision tasks

Another feature of C^3 environments which is missing from most behavioral decision research is their *dynamic* nature. Hogarth (1981) argued persuasively that this omission has led to a number of erroneous conclusions about human' ability to cope in complex environments. Perhaps many of the judgmental biases found in static tasks are irrelevant to dynamic C^3 tasks in which outcome feedback is available.

Two interrelated difficulties are responsible for the failure of most researchers to deal with dynamic tasks. From the experimental side, the difficulty is in inferring clear-cut cause and effect relationships when multiple variables are changing from trial to trial. From the systems modeling side, the difficulty is the intractability of modeling the complex interactions and interdependencies among variables. However, some theoretical progress has been made in this direction (Hall, 1982).

Several lines of behavioral research exist concerning dynamic decision making. Although the findings are quite interesting, adequate mathematical models of the tasks do not yet exist. One crucial issue which arises in dynamic, but not in static tasks is that of whether to seek and assess information before taking action or "shoot first and ask questions later." Kleinmuntz (in press) distinguished between action- and judgment-oriented decision strategies. In a simulated medical decision making task, he found that novices performed poorly because of being too judgment-oriented. In contrast, actual medical practitioners are relatively more action-oriented. This finding is important, because it bears directly on the issue how much effort (workload) is expended on situation assessment versus response selection. If Kleinmuntz' finding generalizes beyond medical decision tasks, one would expect that, whenever practicable, experienced decision makers would rely *more* on trial and error and *less* on thorough situation assessment.

Hogarth and Makridakis (1981) used a complex management simulation game to compare the performance of teams of management students with baselines provided by simplistic rules. They made a number of suggestions concerning how to analyze decision making behavior in a dynamic, competitive environment in which the degree of experimental control is very limited. For example, they compared team performance to the performance of a simplistic but consistent algorithm and a random algorithm.

Lopes and Casey (1987) examined situational influences on risk taking. Subjects played a dynamic game against an opponent or computer. Subjects showed "tactical" shifts in risk taking (i.e., they took less risk when they were near victory and more when they were on the verge of loss), but not "strategic" shifts (i.e., they did not adapt properly to offensive versus defensive roles). Some subjects' risk attitudes were such that they were well-suited for a defensive role, while other subjects risk attitudes were better suited for an offensive role. To the extent that these results are general, the implications for assignment of tasks to individuals in the design of C^3 organizations is clear.

Awareness of phenomena and methods from behavioral decision theory should increase the practical usefulness of the workload surrogate by suggesting what sorts of algorithms are more or less likely to be used by decision makers as a function of certain task characteristics and, perhaps, as a function of individual characteristics such as risk attitude. The discussion now turns to work in cognitive psychology and human performance that may suggest directions in which the workload surrogate could be profitably extended. Specific alternative workload assessment techniques are discussed which may provide benchmarks against which the workload surrogate can be compared in order to identify some of its strengths and weaknesses.

3.2 THE CONCEPT AND MEASUREMENT OF WORKLOAD

This section reviews the current status of cognitive workload as a psychological construct and the resulting implications for accurate measurement of workload. The three major classes of workload measurement techniques (subjective measures, secondary task measures, and psychophysiological methods) are treated. The concept of workload is evaluated from a cognitive psychological perspective. The general approaches embodied in subjective and secondary task measures are then evaluated critically with respect to their psychological and psychometric validity. Finally, practical recommendations for applied workload measurement are made for each of the three classes of measures. The objective is to highlight the state of the art in workload theory and measurement, rather than to provide a comprehensive review. It is hoped that the issues raised herein will facilitate testing and extension of the workload surrogate.

A recent chapter by Gopher and Donchin (1986) provides an extensive review of the concept of workload and its implications for workload measurement. Gopher and Donchin are concerned intimately with the *theoretical* status of the concept. That is, does there actually exist a psychological construct or variable that corresponds to "workload?" Which psychological theories render the existence of this construct plausible? What experimental data are available to allow discrimination between theories under which the construct is valid versus those under which it is not? The discussion that follows has been guided (albeit loosely at some points) by key issues raised by these authors.

3.2.1 Workload as a psychological construct

Evaluation of the psychological underpinnings of the workload concept necessitates consideration of models of information processing from cognitive psychology. Particularly relevant are models that address attentional and short term memory limitations. Such models, especially early ones, have made more or less precise use of the concept of communication channels.

Single channel or "bottleneck" models of attention are highly compatible with the workload concept. According to a single channel model, performance breaks down when the channel's capacity is exceeded. However, evidence for parallel processing, including semantic processing (determination of meaning) of supposedly unattended information, forced researchers to postulate that information is sometimes processed symmetrical in parallel before it reaches consciousness. The contention then is that conscious processing is always carried out serially. The resulting model is one in which there is a mutli-channel to single channel bottleneck which is sometimes located at an early stage in the processing sequence and sometimes located at a later stage. The former situation is referred to as "early selection" and the latter "late selection."

Such a malleable model is of little help in developing a simple concept of workload. Its major contribution is that it points to the importance of task characteristics in determining workload. A task which induces late selection will result in a smaller workload than a comparable task which requires early selection.

One way of determining at what stage of processing filtration will occur in a given situation is to identify whether the operator is using *controlled* or *automatic* processing. The discovery of the controlled versus automatic distinction has been one of the major contributions of cognitive psychology (see Schneider, Dumais, and Shiffrin, 1983, for a review). Two features set the stage for automatic processing: (1) constant mapping between stimuli and responses, and (2) extensive practice in doing the task with this mapping. Automatic processing is done in parallel, is not limited by short term memory, and requires little conscious effort. In contrast, controlled processing is much slower, is limited by short term memory and is relatively effortful. The kinds of decision tasks that are of practical interest tend not to meet completely the consistent mapping requirement of automatically. However, in modeling workload, it may be possible to partition a task into subtasks, some of which permit automatic processing and some of which require controlled processing. In a less fine-grained analysis, the workload of automatic processes could be assumed to be zero. Then, only the subtasks requiring serial controlled processing would need to be modeled.

More recently, attentional limitations have been conceptualized as resulting from demands on multiple, somewhat independent, "resources." According to this view, separate processing resources are available for encoding, deciding (central processing) and responding. Within each of these stages, basically two resources are available. Separate resources are used for encoding visual versus auditory information, processing spatial versus verbal information and for making manual versus vocal responses. In this framework, it is implicit that workload depends in large part upon the degree of *competition* among tasks or subtasks for a single input modality, type of decision operation, stage of processing, or response mode (Wickens, 1984).

3.2.2 Implications for workload measurement

It is clear from the above discussion that the concept of workload as a single, unidimensional psychological quantity is on shaky footing. That is, its *construct validity* is questionable. And, unfortunately, the present state of progress in models of attention and cognitive resource allocation is such that a well-defined, psychologically sound alternative does not yet exist. The next line of attack for the pragmatist in search of a measurement technique having some scientific basis, would be to identify a technique for which *statistical* validity and reliability have been established. However, Gopher and Donchin paint a discouraging picture of existing techniques with respect to these criteria, as well. In the two subsections that follow, two of the general classes of workload measurement techniques, subjective and secondary task measures, are evaluated with respect to their psychological and psychometric validity. Discussion of psychophysiological measures is postponed until the section concerning practical recommendations for workload measurement.

Subjective measures are those which require operators to report, usually via a paper and pencil questionnaire or rating scale, their level of workload. Measures of this type are usually administered immediately after performance of a task. Unfortunately, in terms of theoretical considerations, this family of methods falls short in several respects, including lack of proper psychometric validation, limitations on operators' degree of conscious awareness, limitations on memory of the level of workload, and poor correlation with objective performance.

The typical criterion for acceptability of subjective measures has been that they exhibit "face validity." In other words, the measure is valid if the questionnaire items appear subjectively to be relevant to workload. Psychometrics generally agree that face validity should play some role in the evaluation of a measuring instrument, if only because it impacts directly upon user acceptance. However, objective criteria must be the primary means of evaluation of the instrument. O'Donnell and Eggemeier (1986) offer a practical and balanced set of criteria for choosing between existing workload measures. These criteria are sensitivity, diagnosticity, intrusiveness, implementation requirements and operator acceptance. Of these, only operator acceptance bears a clear connection to face validity. O'Donnell and Eggemeier point out that, as a result of the lack of concern with psychometric theory, there has been little standardization of subjective workload measures.

A potentially more serious problem inherent in subjective measures stems from the fact that most cognitive work is carried on outside the realm of conscious awareness. Presumably, people are not able to report on the workloads of processes of which they are unaware. For example, automatic processing is apparently neither accessible to awareness nor under conscious control; the individual may be aware of the stimulus and usually has control over the overt response, but cannot direct the intervening process. Although subjective measures should be appropriate for tasks calling primarily upon controlled (serial, limited capacity) processes, an experienced operator performing a well-defined task (as is typically the case in tactical battle management) is likely to make extensive use of automatic processing.

A striking feature of the cognitive psychological literature is that consciousness is ascribed a minimal (or no) role in most theories. This reflects the growing realization that awareness is but a small window on the whole of cognitive activity. Unless one is willing to assume that processing limitations are associated uniquely with conscious processing, and that all other cognitive activity is carried out by massively parallel structures, then use of subjective workload measures as the sole source of workload information is inadequate.

Even if the operator is aware of the level of workload *during* the task, additional processing resources are required to *store* this information for retrieval in response to the workload questionnaire. Thus the *retrospective* nature of subjective measures may introduce errors either because the operator has inaccurate memory of the level of workload, or because having to encode workload information changes the workload or the relationship between workload and performance. It is easy to imagine that the accuracy of subjective reports of workload may be quite high when workload is low and plenty of spare processing capacity is available to monitor and store workload information for later recall. However, when the operator is pushed to the limit, accuracy of subjective estimates may drop drastically (perhaps, in the direction of underestimation) due to lack of capacity to process workload information. It has been demonstrated that, when short term memory is heavily loaded, it is possible to encode specific information and yet forget about the entire episode within a few seconds; it takes effort to maintain information in short term memory and additional effort to transfer it to long term memory. If, alternatively, sufficient resources continue to be allocated to workload information despite increased task demands, performance may deteriorate with increasing workload at an artificially low level of workload.

A final shortcoming of subjective measures is one that is not surprising in light of the above arguments: subjective measures are often found to correlate poorly with actual performance. Thus the seemingly trivial assumption that performance will tend to drop as workload is increased from a moderate to a high level is often difficult to confirm when workload is measured subjectively. If subjective measures of this sort are nonetheless assumed to tap something psychologically real, this result may pave the way for designing tasks such that operators perform extremely well, but *feel* that the task is easy. However, the goal of the current organizational design effort is to be able to structure individuals' decision making roles within the organization such that performance (at the organizational level) is acceptable *and* individuals' cognitive resources are not overtaxed. If this goal is to be achieved, having a workload measure which is sensibly related to performance is of paramount importance. This issue is the focus of the Louvet et al. (1988) paper.

Secondary task measures. The rationale underlying the secondary task approach to workload measurement is that increases in the workload of a primary task should be reflected in decreased performance on a concurrent secondary task. Subjects in secondary task experiments are instructed to maintain performance on the primary task at a high level, even if it means sacrificing performance on the secondary task. The assumptions underlying this approach are:

(1) both tasks draw upon a single, fixed "pool" of processing resources (basically, a single fixed capacity channel);

(2) the two tasks compete only for central processing resources, not for peripheral channels -- that is, ability to perform the tasks concurrently is not determined primarily by perceptual or manual limitations;

(3) performance is more or less inversely related to amount of effort

(workload) allocated to the task, so that secondary task performance will vary inversely with the workload of the primary task (assumptions 1 and 2 seem to be necessary, but not sufficient to ensure that this assumption is met);

(4) operators are able to adjust fairly optimally the amount of effort expended on the secondary task as the workload of the primary task is varied; and

(5) the dual task workload is equal to the sum of the workloads for the two tasks when performed separately; that is, the workload of the "meta-processing" necessary to divide effort between the two tasks is insubstantial and the two tasks cannot be "weaved together" in any way to decrease overall workload.

None of these assumptions, other than perhaps (3), are met strictly. Regarding (1), as discussed above, current conceptions of attention implicate multiple pools of resources. it has also been suggested that the pools themselves shrink and expand under certain circumstances. Assumption (2) can be dealt with by attempting to select a secondary task that is complementary to the primary task in terms of its peripheral requirements. However, the need to select a secondary task in this way eliminates the possibility of a single, standard secondary task. Also, largely because of peripheral compatibility considerations, the kinds of tasks that are typically employed in secondary task experiments are rather contrived. For example, in an experiment reported by Gopher, Brickner, and Navon (1982), subjects used a tracking controller in one hand to track a target which moved randomly about a CRT screen. Subjects used the other hand concurrently to make keypress responses to letters superimposed on the target. Assumption (4) is also quite problematic, since dynamic judgments about resource allocation are an explicit part of the task.

To deal with assumption (5), Gopher and Donchin advocate use of a method based on performance operating characteristics (POC). This method is not tied to any particular combination of primary and secondary tasks, so long as the combination meets assumptions (1)-(4). A POC is an empirically derived curve which is a "performance trade-off function that describes the improvement of performance on one task due to added resources released from lowering the standard of performance on another task with which it is time-shared" (p. 41-28). This method, in effect, factors out subjects' inability to set and maintain a certain division of effort between the two tasks. This is a variant of a technique used in many areas of experimental psychology (see Green and Swets, 1966). The underlying theory is generally considered to be quite sound. However, the method requires much more complicated experimental designs and many more observations than other methods of measuring workload. This is because the entire experiment must be repeated a number of times (*at least* four or five) with different trade-off instructions in effect. Gopher and Donchin contend that this extra effort is well spent.

The combination of assumptions (2) and (5) makes development of the design and procedure for secondary task experiments potentially quite tedious. Seemingly subtle and inconsequential aspects of the experimental arrangement may affect whether these assumptions are met and thereby influence in an unpredictable manner the pattern of results. Extensive pilot work is necessary to ensure that these assumptions are met and to be able to counter alternative explanations of the results in terms of assumption violations.

3.2.3 Selection of practical workload measures

In this section practical recommendations are made concerning how to identify the appropriate workload measurement technique for specific applications. These recommendations are based largely on a comprehensive review by O'Donnell and Eggemeier (1986). These authors discuss a number of state-of-the-art methods from the human factors literature. Methods are evaluated with respect to five criteria. These are *sensitivity* (ability to discriminate variations in workload), *diagnosticity* (ability to identify the source of workload or "bottleneck" within the operator), *intrusiveness* (tendency of the workload measure to change the task and workload), *implementation requirements* (instrumentation needed), and *operator acceptance*.

Subjective measures. In general, subjective measures are recommended for ease of implementation (i.e., no specialized apparatus or training is needed), non-intrusiveness (i.e., the measures are typically administered *post hoc)* and, to some extent, sensitivity. The most widely used subjective measure, and the only one which has been subjected to rigorous tests of validity and reliability is the Cooper-Harper scale. This scale was originally developed to measure aircraft ease of handling. However, a more general version of the scale was developed by Wierwille and Casali (1983). This version is applicable to a wide range of systems operation tasks and has been experimentally validated. High correlations with factors affecting objective task difficulty are typical. Notwithstanding the criticisms discussed above that apply categorically to subjective measures, this scale scores favorably in terms of all of O'Donnell and Eggemeier's criteria with the exception of diagnosticity. Subjective measures, including the

Cooper-Harper scale, generally do not permit discrimination between workload due to, for example, central processing versus motor (manual) or perceptual limitations. Therefore, if cognitive workload is the variable of interest, it is essential that the task be designed so that manual workload is relatively small. Operators cannot be relied upon to report only the degree of *cognitive* workload.

In addition to rating scales, a number of *psychophysical measurement techniques* are available for eliciting subjective judgments of workload. These methods are not based on any underlying theory of the nature of human information processing or workload. Rather, they are based on axiomatic measurement theory. This class of methods has a long history in psychophysics and has been used in a myriad of contexts. These methods include:

(1) *Magnitude estimation:* The operator is given a standard or reference task of intermediate difficulty and instructed that the workload associated with this task is, say, 10. Additional tasks, chosen to vary in terms of workload, are then administered and the operator assigns values to these tasks according to the ratio of difficulty of each task to the standard. For example, a task twice as difficult as the standard would receive a value of 20.

(2) *Paired comparisons:* Two tasks are presented serially and the operator is asked to judged which had the higher workload. All possible pairwise combinations of the tasks of interest are presented. The workload for any given task is the proportion of occasions on which it was judged to have the higher workload.

(3) *Conjoint measurement:* This rather intricate technique involves identifying the task attributes and levels that contribute to workload, having operators rate the workload associated with each and every possible combination of attribute levels, analyzing these ratings for consistency with the set of measurement axioms, and, finally, finding a model which accurately predicts workload ratings given the levels of the attributes.

The *subjective workload assessment technique* (SWAT) represents a specific implementation of conjoint measurement. SWAT assumes that three attributes, *time load* (amount of spare time), *mental effort load* (degree of concentration) and *stress load* (strength of feelings of confusion, risk, frustration, anxiety) make up workload. Operators rate tasks on a one to three scale for each of the three attributes. Whenever the conjoint measurement axioms are met, SWAT performs well on all five of the above criteria, with the exception of diagnosticity. SWAT requires specialized software for axiom testing and model fitting.

Unlike scales of the Cooper-Harper variety, techniques based on measurement theory require operators to initially make large numbers of judgments. As a result, these techniques may be impractical outside the laboratory. There is also a risk that subjects will not make considered judgments when so many repetitive and similar judgments are required (Crozier, 1978). Relatively more information may be gleaned from a few carefully considered judgments. Nonetheless, this family of techniques offers the benefit that, if the attributes are chosen correctly and the axioms met, the resulting workload estimates are certain to be valid.

Secondary task measures. A variety of secondary task procedures have been shown to provide valid measures of the workload of the primary tasks with which they have been paired. However, as discussed above, the secondary task approach makes several assumptions about the relation between the primary and secondary tasks. As a result, it seems to be fundamentally impossible to identify a single, universally appropriate secondary task. In the literature, primary and secondary tasks typically come as inseparable packages. In order to fit an appropriate secondary task to a predefined primary task of interest, it is necessary to consider a number of different procedures and identify one which can be *modified* easily to meet the requisite assumptions. Specific secondary tasks suggested by O'Donnell and Eggemeier include tracking, monitoring, memory, mental mathematics, shadowing, simple reaction time, and time estimation. In terms of the five criteria for evaluating workload measures, secondary task measures can be highly sensitive and diagnostic. However, they tend to be quite intrusive, require painstaking implementation and, usually, some specialized apparatus or software. Operator acceptance is of greater concern for secondary task measures than for subjective measures.

Psychophysiological measures. Psychophysiological measures can be classified as relating to either brain, eye, cardiac, or muscle function. Typical measures are *event* [stimulus] *related brain potentials* (ERP) such as *P300, papillary response, heart rate variability,* and *surface electromyographic signals*. A serious difficulty with nearly all psychophysiological measures is that they are sensitive to all sorts of physiological and even psychological variables, many of which do not necessarily correlate with workload. Because of this broad spectrum sensitivity, these methods tend to be low in diagnosticity and sometimes low in sensitivity.

The psychophysiological measure which appears to be the most sensitive and diagnostic is that based on ERP and the P300 component in particular. Gopher and Donchin (1986), in fact, limit their discussion of psychophysiological measures to P300. P300 is measured by processing of the outputs of electrodes attached to the scalp. The P300 amplitude is affected by the degree of task relevance and the degree of unexpectedness of the stimulus (i.e., the stimulus probability). It appears sensitive only to the stimulus evaluation process and not to the response process. The latency of the P300 (time lag between stimulus and P300 signal) reflects the time taken to perceive and evaluate the stimulus.

The usefulness of P300 for measuring workload comes from the finding that the magnitude of the P300 elicited by a secondary task decreases as the difficulty of the primary task increases. Some evidence even exists that the P300 evoked by *primary* task stimuli reflects overall workload. A major advantage of this method is that it is not contaminated by any competition or interference which may occur between the two tasks at the response stage.

A disadvantage of the P300 method is that it seems to require most of the same assumptions as the secondary task method. In addition, the effect of *inwardly* directed attention -- workload in the form of higher order analysis and decision making -- on P300 is not immediately clear. This component of workload can vary somewhat independently of the external attentional demands imposed by stimuli. Loosely put, at issue is whether P300 comes before or after response selection.

In order to ensure low intrusiveness and high operator acceptance of psychophysiological measures, it is necessary that subjects be fully accustomed to the measuring equipment, before actual experimental data are collected. All of the methods require specialized apparatus. Some methods, including the P300 method, require signal processing apparatus and/or software.

Lest the many shortcomings and limitations of the various approaches to workload measurement be taken as overly discouraging, it is essential to bear in mind the broad magnitude and scope of the workload researcher's task. In many applied contexts, the practical benefits to be reaped from a workload measure which accounts for even a *modest* portion of the variance in "true" mental workload are immense. For several of the methods discussed herein, as well as the workload surrogate presented in Chapter 2, this goal appears to be well within reach.

3.3 CONCLUSION

In the context of the ongoing organizational design effort, it is necessary to have the knowledge required to assign tasks to organization members in such a way that full advantage is taken of their information processing abilities without inducing an overload state. Toward this end, the organizational designer must have knowledge of how task characteristics (e.g., time pressure and dynamic evolution of scenarios) affect individuals' information processing and decision making strategies. In addition, large and stable differences in rate (or strategy (e.g., risk seeking/avoiding) of information processing need to be taken into account. A valid measure is needed to provide a quantitative assessment of task workload and what constitutes an information processing overload.

The purpose of this chapter has been to bring to bear on these issues important empirical results and methods from experimental psychology. In the following chapter, an experiment will be described that has been used to evaluate the workload surrogate of Boettcher and Levis (1982) in terms of the validity of its psychological foundations.

4. EXPERIMENTAL METHOD

4.1 INTRODUCTION

In the experimental psychology and behavioral analysis literature, one may find two different approaches which may be related to the concept of human bounded rationality: decisionmaking under time pressure, discussed in Chapter 3, and the Yerkes-Dodson 'law'.

Considerable experimental psychological work has examined the influence of arousal on performance in various types of tasks. Figure 4.1 shows the relationship between arousal and performance called the Yerkes-Dodson 'law'. This relation is shown when arousal is varied over an extremely wide range. Arousal is influenced by a variety of factors including cognitive workload. At very low arousal, performance is low due to boredom and vigilance limitations. At very high arousal, performance is also low, but it is then due to stress and sensory overload. In a well designed organization, all decisionmakers should be operating near the top of the curve.

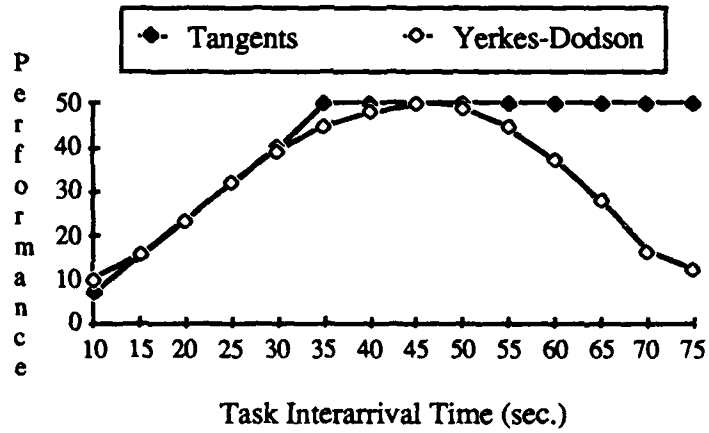


Figure 4.1 The Yerkes-Dodson Law.

Decisonmaking under time pressure, however, has been given very little attention; only a few studies have been reported in the behavioral decision litterature (Ben Zur and Breznitz,1981; Wright, 1974; Wright and Weitz, 1977). The general conclusion is that people under time pressure process only a portion of the information that they would normally process. Further, they filter the information so that what is processed is more important than what is not processed. These conclusions are used as assumptions when modeling the task in the experiment. Time pressure is one of the most significant features of decisionmaking in the context of tactical battle management (Cothier, 1984) .

In the information theoretic model, it is assumed that simple information processing tasks are performed with little error when both the rate of information processing imposed by the input interarrival rate is low and the decisionmaker is not bored. As the input interarrival rate increases, the decisionmaker increases his information processing rate. If the information rate increases further still, a point is reached when the decisionmaker may not increase their processing rate anymore: the decisionmaker is overloaded and his performance decreases significantly. The degradation of performance and the decisionmaker's coping strategies are not statistically predictable and may take many forms. Examples of coping strategies may be ignoring entire inputs, simplifying the algorithms used to give less accurate responses, etc.(Miller, 1969).

The notion that the rationality of a human decisionmaker is bounded has been modeled as a constraint on the total activity G, see Eq. (2.20). Equation (2.20) may be rewritten using the DM's average processing time t as

$$\frac{1}{c_2} \leq F\tau \tag{4.1}$$

For values of τ sufficiently small, noted τ_{min}, the time t required to process the task with acceptable accuracy will equal the lapse of time between two inputs, and the inequality in (2.20) will become an equality described as :

$$G = F_{max}\tau_{min} \tag{4.2}$$

where

$$t_{\text{per input}} = \tau_{min} \tag{4.3}$$

and F_{max} is assumed to be the maximum information processing rate, and t the minimum time required to perform the task with the desired accuracy.

The bounded rationality constraint assumes that if the processing rate F_{max} is exceeded, performance will drop significantly in an unpredictable manner. Equation (4.2) may be rewritten as:

$$F_{max} = G / t_{per\ input} \tag{4.4}$$

where the different quantities have already been described above.

From Equations (4.2) and (4.4), it is apparent that for the purpose of investigating the behaviour of the bounded rationality constraint, the maximum information processing rate is a function of three different parameters: the total activity required to perform the task, noted G, the input signal interarrival time, noted τ, and the minimum time required to process the information and perform the task with the desired level of accuracy, noted t. These conclusions have a significant impact when considering the design of experiments which will be described in the next section.

The existence and the behaviour of the bounded rationality constraint were tested with the experiment described in the next section that was carried out at the MIT Laboratory for Information and Decision Systems. First, the relevant parameters are characterized in section 4.2. Then, the experimental procedures are reviewed in section 4.3. Finally, the purpose of the task constraints and the experimental setup are explained in sections 4.4 and 4.5.

4.2 THE PARAMETER TO MANIPULATE

The information processing rate F has been described in Chapter 2 as being a mathematical function of three different parameters, the cognitive workload required to perform the task, the minimum time required to perform the task for a given level of accuracy, and the input signal interarrival time (see Equations (4.2) and (4.4)). When considering the maximum processing rate F_{max} is considered, these three parameters may be reduced to two, since the assumption is that when F_{max} is reached, the input interarrival rate is equal to the minimum processing rate. As a result, the parameter "time" may be considered as the time allotted to perform the task, also called the window of opportunity. Therefore, two different approaches may be used to study F_{max}. One may manipulate the cognitive workload (G) while the other the time allotted to perform the task (t).

The effect of the bounded rationality on performance as a function of workload or time allotted per trial has been described as a step function (see Figures 4.2 and 4.3.). Performance is stable until the maximum amount of information processing is reached. Then performance drops at or under chance level. The step function represents an instantaneous decrease in performance. It is assumed however, that human decisionmakers will not behave in such a rigid way; when F_{max} is reached, performance will decrease significantly but more smoothly than the step function.

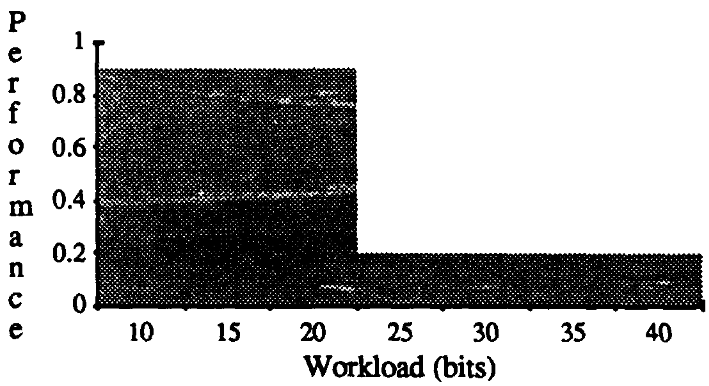


Figure 4.2 Performance as a Function of Workload

The first approach consists of varying the amount of cognitive workload while keeping the time allotted to perform the tasks constant. For a given t, the critical cognitive workload G* associated with F_{max} is measured experimentally as the workload after which performance decreases significantly. The second approach consists of varying the time allotted to perform the task while keeping the workload constant. For a given task, the critical time t* associated with F_{max} is measured experimentally. The total activity G, associated with the task is computed analytically using the information theoretic model.

Manipulation of the task processing time is simpler to monitor and control under experimental conditions than manipulation of workload. In particular, time is a continuous variable whereas the workload associated with different tasks takes discrete values and needs to be assessed analytically. Therefore, the time allotted per trial was the manipulated parameter.

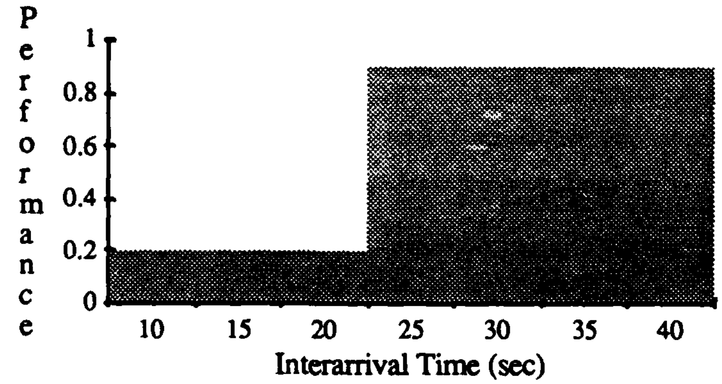


Figure 4.3 Performance as a Function of Interarrival Time

4.3 EXPERIMENTAL PROCEDURE

This work is only the first in a series of experiments, therefore the simplest decision making organization was simulated: the organization was reduced to a single decision maker. Since little was known about the experimental aspects of estimating the bounded rationality constraint, the task was designed so that the factors which were affecting the subjects' performance could be monitored as precisely as possible. The task was also chosen so that the subjects could become 'well trained experts' with reasonable amount of training, thereby satisfying the requirement that the decisionmakers' performance did not benefit from the learning effect during the experiment.

4.3.1 Experimental Conditions

The experiment consisted of a highly simplified tactical air defense task . It was run on a Compaq Deskpro Model 2 equipped with an 8087 math coprocessor, monochrome graphics card (640 X 200 pixels), 640K of memory, and monochrome monitor. The experiment was programmed in Turbo Pascal version 3.01A. The operating system was MS-DOS version 2.11. It was also run on an IBM PC AT with the 80287 math coprocessor and with 640K of memory. None of the high resolution graphics capabilities of the AT were used so that the experiment be portable to a wide variety of PC compatible machines.

The computer screen shown in Figure 4.4 consists of three different parts: A large circle, a small circle and a rectangular box. The large circle represents a radar screen. The small circle represents the clock which shows the time allotted for the trial as well as the amount of time left to perform the task. The rectangular box on the left of the screen and full of 'domino ' shaped rectangles, shows the number of ratios used for the given trial (four in this example) and the number of ratios still to be processed (two in this case). The keyboard was used to enter the subjects' responses.

The experiment consisted of blocks of trials. A trial consisted of either four or seven threats that were to be processed by the decisionmaker within the allotted time shown by the clock. Within each block of trials, the number of ratios was constant and the time allotted per trial was varied in alternating descending and ascending order. Each block of trials was seperated by a longer pause and flashing to indicate that the number of ratios was changing.

For each threat, two pieces of information were presented as a ratio of two two-digit integers: relative speed and relative distance from the center of the screen. The distance was in the numerator and the speed in the denominator. Therefore, each ratio represented the time it would take the threat to reach the center of the screen. The subject's task was to select the threat which would arrive first at the center of the circle in the absence of interception. The task can be interpreted as one of selecting the minimum ratio.

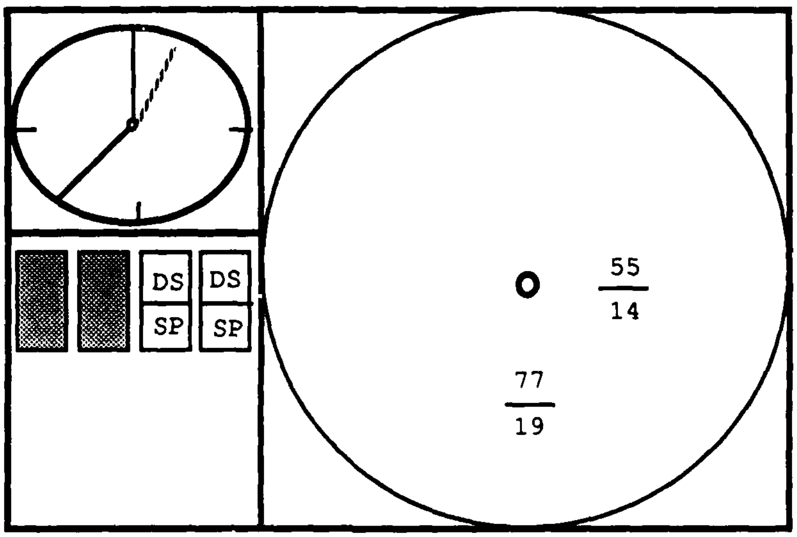


Figure 4.4 The Screen Display Used in the Experiment

For each trial, only two ratios were identifiable and present on the radar screen at the same time. The other ratios were shown on the side of the screen by the 'domino' shaped rectangles. Such a procedure forced the DM's to process ratios in pairs.

The ratios appeared only on the vertical or horizontal diameter of the radar screen, and the physical distance of each ratio from the center was proportional to the distance of the ratio as indicated by the numerator. Thus ratios appeared in one of four regions: left, right, above, or below the center. Each ratio was randomly assigned to one of these four regions, subject to the constraint that no two ratios appeared in the same region at the same time. For each pair of ratios in a given trial, the subject indicated his or her choice by pressing one of four arrow keys corresponding to the direction of the ratio from the radar screen's center. The ratio chosen as smallest was retained on the radar screen, the other vanished, and the next ratio to be processed was taken from the small rectangle's area and placed on the radar screen. This procedure was repeated until all ratios of the trial had been examined. Row(s) of small rectangles to the left of the radar screen indicated the total number of ratios for the current trial and the number yet to be examined (see Figure 4.4). Each time a new ratio appeared on the radar screen, one of the rectangles turned grey and the numbers within that rectangle disappeared. The subject could not give a final answer until all the ratios had been examined, (three comparisons for four ratios and six for seven). The arrow keys were located on the numeric keypad of the keyboard and were arranged isomorphically with the four regions of the radar screen.

Performance feedback was provided at the end of the trial. When a trial was finished on time, only one ratio remained on the screen at the end of the trial. If this ratio was in fact the smallest, it "flashed" several times to indicate a correct response. If this ratio was not the smallest, a low-pitched tone signalled the error. This tone (which subjects reported to be particularly obnoxious) was used to discourage subjects to use guessing as a strategy. When a trial was not finished on time, the screen vanished so the subject knew he had not answered within the allotted time.

4.3.2 Manipulation of Task Interarrival Time

In usual information theoretic setups, it is assumed that the inputs are emitted by one or many source(s) at a mean symbol interarrival time noted τ. In this experiment, to test the existence of the bounded rationality constraint, the average interarrival time is not held constant, but is varied. However for easier control of the experimental parameters, the time allotted to perform the task (noted t) is monitored, not the interarrival time.

The amount of time allotted for each trial was shown by the fixed clock hand (see Figure 4.4). A moving second hand (running clockwise from twelve o'clock) indicated elapsed time within a trial. A one and a half second pause prior to the start of each trial allowed subjects to see how much time was allotted. The fixed hand flashed during this interval. Time allotted per trial was varied in alternating descending and ascending series.

One of the questions which were to be answered by this experiment related to the stability of F_{max} across tasks, if it could be shown that F_{max} existed. Two different numbers of ratios were selected to investigate this issue. Therefore one of the questions was

$$\frac{G(4)}{t^*(4)} \overset{?}{=} \frac{G(7)}{t^*(7)} \tag{4.5}$$

This issue raised another question: When considering the measurements of time allotted per trial, should the time allotted per trial be considered or should the average time allotted per comparison for each trial be considered?

One of the hypotheses was that because of the task setup which only allowed the subjects to consider two ratios at the same time, the cognitive workload required to process the four ratios was approximately twice that required to process trials of seven ratios. In one case three comparisons were required whereas in the other six comparisons were required, and it was assumed that the same algorithmic structure was repeated for each comparison. Equation (4.6) shows the workload for one comparison, whereas Equation (4.7) shows it for two comparisons.

$$G_1 = H(x_1) + \sum_{i=1}^{k} H(w_i) + H(y_1) \tag{4.6}$$

$$G_2 = H(x_2) + \sum_{i=1}^{k} H(w_i) + \sum_{i=k+1}^{2k+1} H(w_i) + H(y_2) \tag{4.7}$$

where x_1 is the input variable and y_1 the output variable for one comparison, and x_2 is the input variable and y_2 the output variable for two comparisons, and there are k internal variables noted w_i for each comparison.

Assuming that the workload per comparison was approximately the same for four and for seven ratios, if it were proved experimentally that the minimum average time allotted per comparison was not significantly different for four and for seven ratios, then F_{max} for both numbers of ratios should be assumed to be not significantly different. Therefore, it was decided that the parameter which should be monitored was the average time allotted per comparison which will be noted T, rather than the time allotted per trial which was noted t. T may be expressed as a function of the number of comparisons m within a given trial as follows:

$$T = t/m = t/n\text{-}1 \qquad (4.8)$$

where n is the number of ratios. To study the variations between trials of three and trials of six comparisons, the average time per comparison was set to be the same for both types of trials. (Assuming F_{max} exists, the time threshold associated with F_{max} would be derived from the experimental results, and noted T^*_3 for three comparisons and T^*_6 for six.)

The experiment was also constructed to minimize the influence on performance of time required for non-cognitive (i.e., perceptual and motor) activity. A trial consisted of a set of either three or six comparisons. For a set of three comparisons, the time allotted per trial, noted t, ranged from 2.25 to 10.5 seconds. For a set of six comparisons, t ranged from 4.5 to 21 seconds . Thus the average time per comparison, noted T, was varied from 0.75 to 3.5 seconds in 0.25 seconds increments for both conditions and 12 different values of T were recorded. Since even the minimum average time per comparison of 0.75 seconds allowed ample time for eye movements, perception, and motor response, it could be assumed that the major limiting factor on the performance of the subjects was the bounded rationality constraint F_{max}.

4.3.3 Organization of Trials

The experiment consisted of blocks of twenty four trials within which the number of ratios was kept constant. A block of trials consisted of a descending series over the 12 values of t, followed by an ascending series. Such an alternation between ascending and descending series was aimed at smoothing out the anchoring effect of either only going from minimum to maximum or only going from maximum to minimum. After a block was over, the number of ratios was changed for the subsequent block. There was a 2.5 sec. pause between blocks, during which time, the large rectangle to the left of the radar screen (see Figure 4.4) flashed to indicate the impending change in the number of ratios. The pause was aimed not only at showing to the subjects what the next number of ratios would be, but also at reducing tension.

For each subject, the full experiment consisted of eight blocks of trials for both numbers of comparisons. The number of comparisons changed at the end of each block. The small differences between the difficulty of different trials were to even out when considering blocks of twenty four trials. The subject's response was recorded and mapped with the expected solution. Immediate feedback showed the subject whether the answer was correct or not. Such a method satisfied the subject's curiosity about the accuracy of his previous decision. It also allowed the experimenter to estimate the subject's overall performance and ability to cope with time pressure.

The goal was to study the subjects' degradation of performance. Therefore it was important to make sure that the range of time intervals for which the subjects were tested was large enough so that both a stable performance and a degradation of performance could be observed. The subjects had to be tested both over time intervals that were large enough so that their performance was close to optimum, and also small enough so that their performance be below chance level.

By observing the subject run one session of the experiment, it could often be estimated if the experiment was well calibrated for the particular subject, i.e., if the time window used to test the subject was well chosen. For some of the subjects the experiment was run over larger time intervals because preliminary analysis of their data showed that the time window used was not large enough to gather all the relevant information. Since an inappropriate experimental setup was not always spotted on time, subjects for whom the experiment was not run properly were asked to come for extra sessions. As a result, for some subjects, more data has been collected.

For the subjects who only came for the scheduled sessions, the total duration of the experiment was approximately 2.5 hours, divided in three sessions: eight blocks of twenty-four trials were completed in each session and subjects typically participated in no more than one session per day. To limit fatigue, each session was seperated into four ten-minute subsessions between which the subjects could take a brake. This was to allow them to relax and have good attention span during the short subsessions. Prior to each experimental session, subjects were given a brief (three to five minute) "warmup" period during which no data were recorded.

4.3.4 Practice Session

Subjects received a 30 minute practice session prior to the actual experiment. This session consisted of six blocks of trials over T for each number of ratios. For the practice session, T was varied from 1 to 5 sec. per comparison in 0.5 sec. increments. Informal discussion with subjects indicated that most felt their performance would not improve substantially with additional practice. Practice was important because the subjects were not supposed to improve their performance as the experiment was run; the analytical tools developed by Boettcher et al. assume that the subjects are both well trained and qualified to perform the task. The practice session was also useful in getting some feedback from the subjects. A few subjects decided not to go on with the experiment, whereas some were advised not to participate in the study. The few subjects who were asked not to participate were people who were not familiar at all with approximation or rounding-off procedures necessary for such a task. They could not meet one of the requirements necessary to use information theory when applied to decision making or decisionmaking organizations: well trained and qualified decsionmakers. Except for those few special cases, the practice data were not analyzed.

4.3.5 Subjects

Twenty-five subjects ran the experiment to its full extent, since one subject was eliminated from the sample. Almost three quarters of the subjects (nineteen) were present or former MIT students (both graduates and undergraduates), the others were MIT employees or students' friends. The large number of MIT students is not inappropriate since MIT students should be qualified to perform the task and, as mentioned above, the subjects should satisfy this requirement.

4.4 PURPOSE OF VARYING THE NUMBER OF RATIOS

It was assumed in section 4.3.2 that the amount of workload per comparison was approximately the same for trials of four and seven ratios. However, the effect of manipulating the number of ratios was of some intrinsic interest, because of implications for how subjects manage their time. Effective time management is more critical for seven than for four ratios, while "overhead" or "start-up" time is more critical for four ratios than for seven.

Therefore, if the value of the subjects' threshold (assuming it exists) was not significantly affected by changes in the number of ratios, it could be established that, to some degree, the bounded rationality constraint is stable across tasks. If, however, instability were found for such a minor task change, there would be no need to go further.

Subjects knew before the start of each trial how much time, t, was allocated for the trial. Part of the subject's task was to budget the available time over the three or six comparisons so that all comparisons could be completed and full use made of the available time. The criticality of accurate budgeting can be seen from Equation (4.9).

$$\text{Response Time} = m\,T' + b \tag{4.9}$$

where m is the number of comparisons (three or six), T' is the average amount of time the subject allocates to each comparison, and b is the overhead, startup, or initialization time for a trial. It is assumed that the value of b is independent of m. According to this model, the subject must choose T' so that the resulting response time is less than or equal to t. Clearly, with increasing m, the detrimental effect of setting T' non-optimally increases relative to the detrimental effect of the fixed overhead, b.

4.5 PURPOSE OF THE TASK CONSTRAINTS

4.5.1 Constraints on the Experimental Setup

In order to constraint the strategies the subjects could use, two restrictions (already mentioned in section 4.4) were imposed. First, ratios were displayed in pairs and only one pair was identifiable at a time. Second, a final response was permitted only after all of the four or seven ratios had been displayed. These two procedures forced the subjects to make a given number of comparisons -three when four ratios and six when seven- or at least forced them to consider all the ratios. Having a more precise idea of the steps the subjects went through is an essential tool when computing the workload, since workload is dependent on the amount of information that the subjects process. Such restrictions also eliminated the variation in the order of information acquisition which could increase the workload, if the subjects had been hesitant when deciding which ratios to consider first.

Within the rest of the thesis, since one of the goals is to study the difference between trials of three tasks and trials of six tasks, a trial will be defined as a set of three or six tasks, where one task corresponds to finiding the smallest of two ratios.

4.5.2 Instruction to the Subjects

Subjects were instructed to attend only to the numeric information of each ratio even though the physical distance of each ratio from the center was proportional to its numeric distance. This was done to restrict the number of strategies the subjects would use.

This restriction is important, because Greitzer and Hershman (1984) showed that an experienced Air Intercept Controller tended to use physical distance information only (and not speed information) in determining which of a number of incoming ratios to prosecute first. This simplified strategy was labeled the *range* strategy. The operator was, however, able to use both range and speed information -- the *threat* strategy -- when instructed explicitly to do so. The threat strategy, if executed in a timely way, is of course more effective than the simpler range strategy.

4.5.3 Constraints on the Ratios

Another method, which was used to monitor as closely as possible the amount of work the subjects did, was to impose constraints on the ratios. The ratios were very carefully chosen to equalize the difficulty of the different comparisons and trials. (Changes in performance were not to be caused by differences in task difficulty, but because of overload.)

For each trial, all ratios were either greater than or less than one. This restriction was included because pilot work had shown that decisions involving ratios on opposite sides of one were trivially easy, regardless of interarrival times. The greater-than-one / less-than-one determination was made randomly for each trial.

Speeds and distances were selected subject to the following constraints:

(1) greater than 10 and less than 98,

(2) no multiples of 10.

(3) Each speed and distance combination was screened and rejected if the resulting ratio was a whole number,

Additional constraints were that :

(4) no speed value be used more than once per trial;
(5) no distance value be used more than once per trial;
(6) no speed value be the same as its corresponding distance value; and
(7) no two ratios have the same value.

Distances were selected independently of speeds, but had to satisfy constraints six and seven.

The second round of pilot experiments included these constraints. The subjects, however, reported that some comparisons were still much easier than others. It appeared that the ratios less than one could be very difficult to compare because the numerical values could be very close. To avoid especially difficult comparisons, new constraints were imposed on trials . As a result, the candidate ratios obtained applying all the constraints mentioned above were screened against the following new criteria:

(8) each possible pair of ratios within a trial of ratios less than one must differ by no less than 0.05 and by no more than 0.9 and;
(9) in the greater than one condition, the minimum allowable ratio was 1.2;

If a candidate ratio failed on any criterion, a new ratio was generated and the process was repeated until a complete set of four or seven compatible ratios had been obtained. (An attempt was made to impose the same constraints on both the ratios less than and larger than one, but when doing so, it was sometimes impossible to generate seven ratios larger than one satisfying the appropriate constraints.)

4.6 FEEDBACK FROM THE SUBJECTS

Generally, subjects seemed to be challenged by the experiment. Many subjects reported that the experiment forced them to concentrate hard and that they were glad that each session was seperated into subsessions between which they could relax. Also, it was a common feeling that there was a breakpoint after which they could not process the task within the required time anymore. A few subjects mentioned that they had had a harder time with trials consisting of ratios larger than one than with ratios less than one. Such a difference was not built in purposely, but is described and explained in Chapter 7; the algorithms which were used by the subjects resulted in a higher performance for ratios larger than one than for the ones less than one. Also, some subjects reported having a difficult time with the keyboard: the response that they had chosen was not always the response that they entered through the keyboard. (Most of the subjects made at least one error just because they had just hit the wrong key!) Such errors will be one of the sources of noise and discrepancies which are found in the data. Finally, it appeared that there was a delay between the instant when the key was pushed and the answer was recorded. This delay was particularly critical for the small values of T, since subjects tended to answer as late as possible; sometimes their right answer was not recorded.

5. EXPERIMENTAL RESULTS

In Chapter 4, the experimental setup was described. In this chapter, the experimental results are analyzed with respect to the hypotheses that may be tested experimentally. First, in section 5.1, the data recorded during the experiment are presented and the hypotheses are stated. In section 5.2, the methodology used to test the different hypotheses is described. In section 5.3, the procedures required prior to testing the hypotheses are presented. In section 5.4, the data are analyzed according to the different procedures and, in section 5.5, conclusions are drawn from the experimental results.

5.1 THE DATA AND THE HYPOTHESES

5.1.1 The Data Collected

This section first describes the recorded measurements and then two examples are given to explain how to reconstruct the data from the recorded data files.

For each trial, seven different data sets were recorded. (See Table 5.1) First the average time allotted per task is shown in column 1. The average time varied between 0.75 sec. to 3.5 sec. for most subjects. The number of ratios for the trial is shown in column 2: either four or seven ratios, i.e., three or six tasks. In column 3 is noted whether the time per trial was increasing or decreasing: 1 indicates a descending series whereas 2 indicates an ascending series. The subjects' performance is recorded in column 4. The subjects received a score of 0 if an answer was given but did not match the correct answer, a score of 2 if no answer was given within the allotted time, and finally a score of 1 if the answer matched the correct one. Column 5 lists the two digit distance, followed by the two digit speed of each ratio in the order it appeared on the radar screen. In column 6 are inscribed the ratio number that the subject chose at the end of each comparison. Finally in column 7, the time (in hundredths of a second) the subject used to process each task is noted.

As an example of how to read the data files, two rows of Table 5.1, (noted *1 and *2 in the table), are described. The trial recorded in the row, *1, may be described as follows. The average time T per task was 3.00 seconds, and there were four ratios, (three tasks), in this trial (as indicated by the 4 in column 2). Then, the 1 in column 3, indicates that this trial is part of

Table 5.1 *Sample of the Data Collected:* Subject 50, Session 1, First Set of Three Tasks

	Col. 1 Time T	Col. 2 # of Ratios	Col. 3 Asc./ Desc.	Col. 4 Perf. J	Col. 5 Speed and Distance of the Ratios	Col. 6 Result of Comparison			Col. 7 Elapsed Time to Completion of Task #		
						1	2	3	1	2	3
	3.50	4	1	1	2686316766873891	1	1	1	204	99	127
	3.25	4	1	1	7344513949248857	2	2	2	214	308	290
*1	3.00	4	1	1	4364185844521563	2	2	4	181	110	165
	2.75	4	1	2	5919652537139531	2	3	3	368	247	220
	2.50	4	1	1	8297298431424676	2	2	2	241	71	82
	2.25	4	1	1	1289368253656283	1	1	1	132	77	55
	2.00	4	1	1	4652118619514157	2	2	2	104	104	49
	1.75	4	1	2	3764111562971634	1	1	0	373	161	0
	1.50	4	1	1	3161179212425881	2	2	2	176	66	38
	1.25	4	1	2	5716822144129622	1	0	0	395	0	0
	1.00	4	1	2	2769347114634358	1	0	0	296	0	0
	0.75	4	1	2	7139763588657537	1	0	0	242	0	0
*2	0.75	4	2	2	6245934837228267	1	1	0	192	11	0
	1.00	4	2	2	3192218148724351	1	0	0	302	0	0
	1.25	4	2	2	6947743525166452	1	1	0	302	82	0
	1.50	4	2	2	7596488753865563	2	2	0	201	230	0
	1.75	4	2	1	1452139539692939	2	2	2	182	55	44
	2.00	4	2	1	2555146124311798	2	2	4	181	104	151
	2.25	4	2	1	5369164165752785	2	2	4	307	127	137
	2.50	4	2	1	2233269464752959	2	2	2	187	99	105
	2.75	4	2	1	4383647834393763	1	1	1	242	131	104
	3.00	4	2	0	5691135651926887	1	3	3	126	225	132
	3.25	4	2	1	9862685588489673	2	2	2	263	121	263
	3.50	4	2	1	2779596614213681	1	1	1	159	94	258

the descending series: the T value was larger before this trial. The 1 in column 4 indicates that at the end of the trial, the subject had correctly chosen the smallest of the four ratios.From column 5, the value of each ratio for this particular trial may be read. The four different ratios were:

$$R_1 = 43 / 64 \quad R_2 = 18 / 58 \quad R_3 = 44 / 52 \quad R_4 = 15 / 63$$

From columns 6 and 7, the following information may be derived. Subject # 50 used 1.81 seconds (column 7, first number) to decide which was the smallest ratio of the first task: The ratio # 2 was chosen, (see column 6, first digit). Then, between the result of the first task and that of the second, 1.10 seconds had elapsed (see column 7, second number), and the subject had chosen ratio 2, (see column 6, 2nd digit). Finally, it took the subject 1.27 seconds to compare the last two ratios (ratios 2 and 4), and enter the final solution, ratio 4.

The trial recorded in the row, *2, may be described as follows. There were four ratios, (three tasks), and the average time per task was 0.75 seconds. This trial was during an ascending series (a 2 in column 3), and the subject did not answer in time, (indicated by a 2 in column 4). The values of the four ratios were as follows, (see column 5):

$$R_1 = 62 / 45 \quad R_2 = 93 / 48 \quad R_3 = 37 / 22 \quad R_4 = 82 / 67$$

Finally, the subject chose ratio 1 as the smallest of ratios 1 and 2 after 1.92 sec. and ratio 1 again as the smallest of ratios 1 and 3 after 0.11 sec. The subject then ran out of time before entering a final solution.

5.1.2 The Hypotheses

The hypotheses which were to be tested using the experimental results were the following:

Hypothesis(1): Decisonmakers are subject to the bounded rationality constraint, that is the bounded rationality constraint sets an upper limit on the amount of information that decisionmakers can process before their performance decreases drastically.

Hypothesis(2): If the bounded rationality constraint exists, assuming that the workload for six tasks is approximately twice that for three tasks, (see section 5.2), is there a significant difference between the value of the bounded rationality for three tasks and that for six tasks for each subject?

In Chapter 8, two more hypotheses are tested combining the experimental and analytical results. The first is designed to confirm that F_{max} is stable for each subject as the number of tasks is varied. The second tests the stability of F_{max} across subjects.

5.2 THE PROCEDURES TO TEST THE HYPOTHESES

5.2.1 The Existence of the Bounded Rationality Constraint

This section first describes the tests necessary to prove the existence of the bounded rationality constraint. Then, the theoretical model, 'single step ', and the empirical model, growth curve, are discussed. Finally, the growth curve is characterized.

In section 4.2, the theoretical model associated with the existence of the bounded rationality constraint is described as a 'single step' function. Such a model is not feasible when considering concrete applications; humans do not behave in such a rigid and structured way, and unwanted noise always distorts experimental results. The empirical model which will be used to prove the existence of the bounded rationality constraint is a growth model (described in the next paragraph). The first hypothesis, the existence of the bounded rationality constraint, may be restated in terms of growth curves as follows:

(1) a growth model fits the data well;
(2) a growth model will fit the data better than a linear model;
(3) the existence of a time threshold (which will be noted T*), may be identified and constructed from the growth curve model. This threshold corresponds to the corner point of the step function shown in the theoretical model Figure 4.2.

The existence of F_{max} will be proved first by showing that the growth curve is a good model of the data, i.e., it has the same general characteristics and a large R^2. The second step will be to show that a growth curve fits the data better than a straight line, i.e., it is possible to identify a time threshold (breakpoint) after which performance decreases significantly. This will be done by showing that R^2, the coefficient of multiple determination, is consistently larger for a growth curve fit than for a linear fit. (In a third step, the time threshold T* is evaluated for each subject in section 5.4.3)

The following paragraphs describe the general attributes of the family of growth curves.

These curves are characterized by an S shape: the growth starts slowly (characterized by a nearly flat curve segment), then the growth increases rapidly (steep slope) and finally levels off. A growth curve seems most appropriate to describe the experimental data, since it characterizes patterns where quantities increase from near zero to close to the maximum level very rapidly.

For the purpose of this experiment, the most appropriate curve of the family of S curves is the Gompertzt curve which has the characteristic of not being symmetric about the inflection point. This is a relevant property, since one can not predict that performance will decrease in a symmetric way after the subject is working beyond the bounded rationality constraint.

The Gompertz curve has three degrees of freedom and is given by (Martino, 1972):

$$J(t) = a\, e^{-b\, e^{-ct}} \tag{5.1}$$

where J is performance expressed as a value between 0 and 1. The Gompertz curve may be characterized the following way: The asymptotes are:

$$\text{At } t = 0,\quad J(0) = a\, e^{-b} \tag{5.2}$$

$$\lim_{t \to \infty} J(t) = a \tag{5.3}$$

The inflection point occurs at :

$$t_{inf} = \ln(b) / c \tag{5.4}$$

and the value of J at the inflection point is:

$$J_{inf} = a / e^{1} \tag{5.5}$$

For linear regression using the least squares method, the Gompertz function may be linearized as follows:

$$Y = A\,X + B \tag{5.6}$$

where

$$Y = Ln\,(Ln\,(a/J))\,;\quad X = t\,;\quad A = -c\,B;\quad B = Ln(b) \tag{5.7}$$

5.2.2 Stability of F_{max} Across Similar Tasks

When considering the experimental results, the stability of F_{max} may be studied assuming that the workload for six tasks is approximately twice that for three tasks. (See section 5.4) Therefore, in this chapter, the stability of F_{max} is tested only with respect to T*, the time threshold (introduced in sections 4.2 and 5.2.1). T* is assessed for each subject for both three and six tasks in section 5.5, after the existence of the bounded rationality constraint has been proved. Then, the distribution over subjects of T* for three and six tasks is evaluated separately, and the type of each distribution is compared. Finally, the significance of the difference between the mean of the T^*_3 and T^*_6 distributions are compared using a statistical test, the t test. The hypothesis is validated, if the statistical tests conclude that the two distributions are of the same type and the means are not significantly different. (A 0.95 level of confidence is used.)

5.3 THE PROCEDURES PRIOR TO TESTING THE HYPOTHESES

5.3.1 The Data Analyzed

Since the hypotheses focused on the subjects' performance, only the data strictly related to the subjects' performanc -- the time alloted per trial, the number of ratios for the given trial and the score for the given trial -- are analyzed. (The rest of the data could provide basic data for future research.)

When assessing performance, a wrong answer and an incomplete answer were treated similarily. As a result, for subject i, for each trial k corresponding to the average time T_j, the score was assumed to be an independant Bernoulli variable with probability $p_{ij.}$

$$X_{ijk} = \begin{cases} 1 & \text{If the tasks were completed within the alloted time and the correct ratio was chosen.} \\ 0 & \text{Otherwise.} \end{cases} \tag{5.8}$$

An estimate of $p_{ij,}$ was computed as follows using the simple unbiased estimator $\bar{p}_{ij}$

$$\bar{p}_{ij} = \sum_{k=1}^{24} X_{ijk} \,/\, N_o \tag{5.9}$$

where N_O is the number of times the subject was run for each time interval. For most subjects N_O is equal to 24. The estimated performance was plotted against the average time allotted per task for Subject #23 in Figure 5.1 and in Figure 5.2 for Subject # 35.

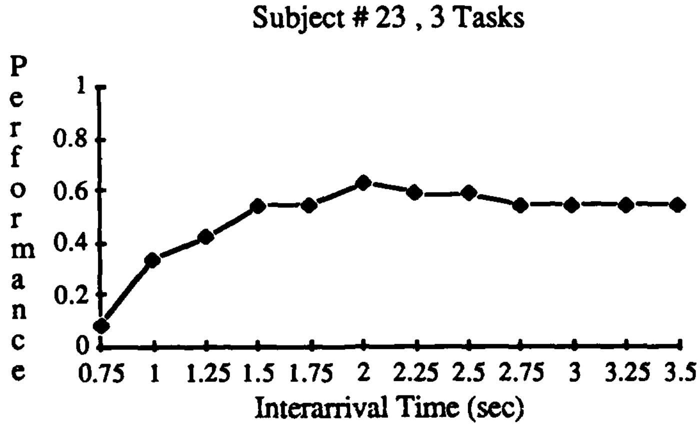


Figure 5.1 Performance Versus Average Allotted Time

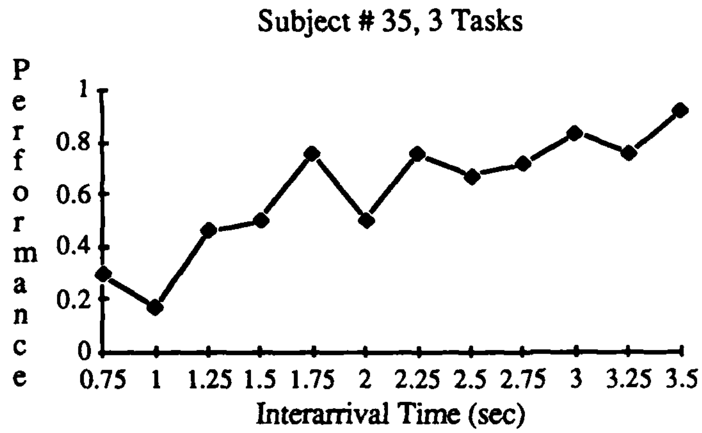


Figure 5.2 Performance Versus Average Allotted Time

5.3.2 Data Transformation

Curve fitting is used to test whether the Gompertz model fits the data well. Since each p_{ij} is the sum of N_O independent identically distibuted Bernoulli variables divided by N_O, each p_{ij} has a different error variance, and one of the necessary assumptions for regression and curve fitting, i.e., equal error variances, is violated.

$$\text{Variance}(p_{ij}) = p_{ij} * (1 - p_{ij}) / N_o \qquad (5.10)$$

Therefore, in order to equate the error variances, the estimates p_{ij} were transformed using the arcsine formula:

$$(\sin^{-1}(\mathrm{sqrt}(\ p_{ij}))) / 1.57 \qquad (5.11)$$

The denominator ($\pi/2$) is a scaling constant to keep the range of the estimates between 0 and 1; the variances remain equal. The arcsine transformation was used instead of the logit transformation because the logit transformation is more appropriate for data which is symmetric about an inflection point. Table 5.2 shows the impact of the arcsine transformation on seven different values ranging between 0 and 1. (Values 1/4, and 1/7 have been chosen since they are the performance which would be expected if the subjects were simply guessing for the trials of three and six tasks respectively.) The general effect of the arcsine transformation is to increase slightly small values, while slightly decreasing large values. Since it has most effect on both the lower and upper values, the arcsine transformation will tend to make a threshold, (if there is any), less visible. The difference between maximum and minimum performance is reduced as the whole curve is 'squeezed' and flattened.

All analyses reported herein are based on the transformed estimates which will be called performance.

Table 5.2 The Effect of the Arcsine Transformation

Value	Transformed Value
0.0	0.0
1/7	0.247
1/4	0.334
0.4	0.436
0.5	0.500
0.6	0.564
0.8	0.705
0.9	0.796
1.0	1.000

5.3.3 The Gompertz Curve Regression

A computer package, RS/1, (Bell Labs) was used to estimate the Gompertz curve parameters for each data set, and evaluate the fit, the R^2. The program first asked for the function to use as a curve fit. The Gompertz function was typed in. Then it asked where to find the x values and the y values; these were stored in a table, the same for all subjects. The program then wrote the partial derivative of J with respect to a, b and c, and asked for starting values for a, b and c, as well as a convergence criterion. The selected starting value for a was different for each subject since the subjects' maximum performance was chosen. The same starting values for b and c were entered for every subject, 2 for b and 1 for c. Choosing different starting values in the same range would not have made any significant difference since for each subject the program ran by iteration until the error converged was less than 0.0001. When a performance of 0 was encountered, the computer transformed it to a small value, apparently in the range of 0.00001.

5.4 APPLICATION OF PROCEDURES AND RESULTS

5.4.1 General Characteristics of the Data Analyzed

Performance versus average time allotted per task was plotted for each subject for both three and six tasks for the transformed data. The curves appeared to have the following set of characteristics:

(1) They do not have the Yerkes-Dodson concave shape. This indicates that the experiment succeeded in tapping into the moderate-to-high arousal portion of the Yerkes-Dodson curve (see Figure 4.2), rather than the "vigilance" portion.

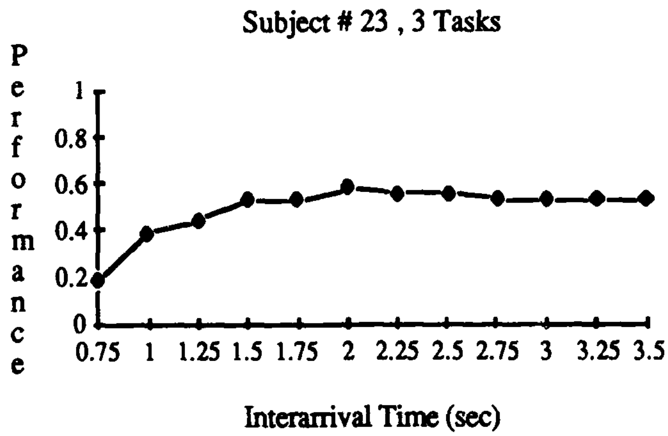
(2) Most curves tend to be flat (zero slope) for large values of T.

(3) They have positive slopes for smaller values of T.

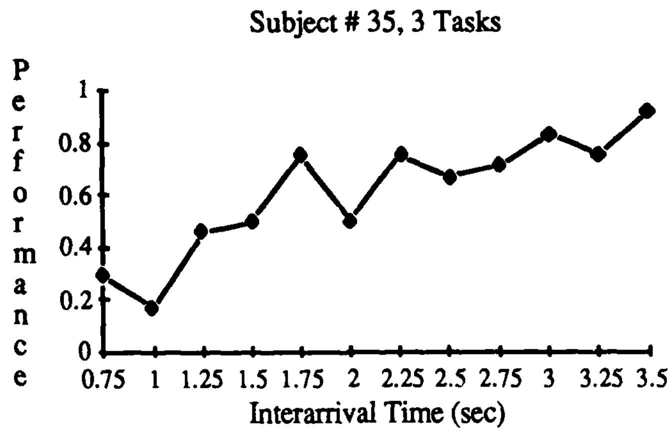
(4) Performance drops and tends to level off for small values of T.

Figure 5.3 shows performance versus the average time allotted per task, t, for two subjects. These curves were selected as being examples of strong, (a), and average, (b), representation of the threshold hypothesis. (These curves are the same as in Figures 5.1 and 5.2, but with the estimated performance.)

Only half of the subjects had more than one data point below chance level because the allotted time could not be decreased indefinitely. It was necessary that poor performance be caused by mental and not physical limitations. The subject needed enough time to press a key. One subject was eliminated from the sample, because the experiment was not run properly (inappropriate time window) and the subject was not available for further testing. As a result, the population sample was reduced to twenty-five subjects.



(a)



(b)

Fig.5.3 Transformed Performance versus Average Allotted Time per Task for Two Subjects.

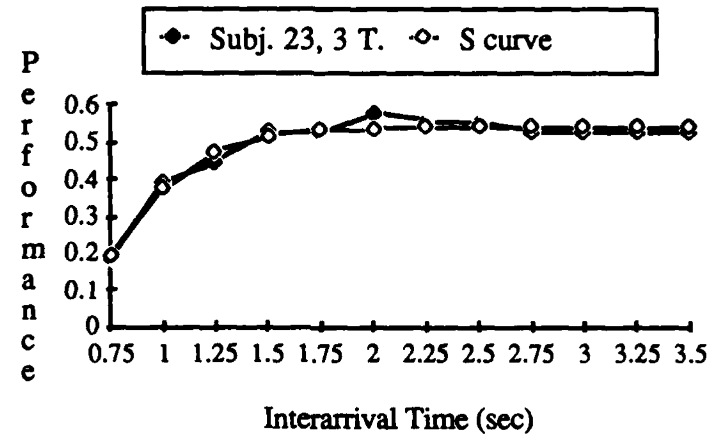
The characteristics of the curves describing subjects' performance as a function of average time allotted per task, suggest that a Gompertz curve could be appropriate for summarizing the data.

5.4.2 The Existence of F_{max} : the Gompertz Fit

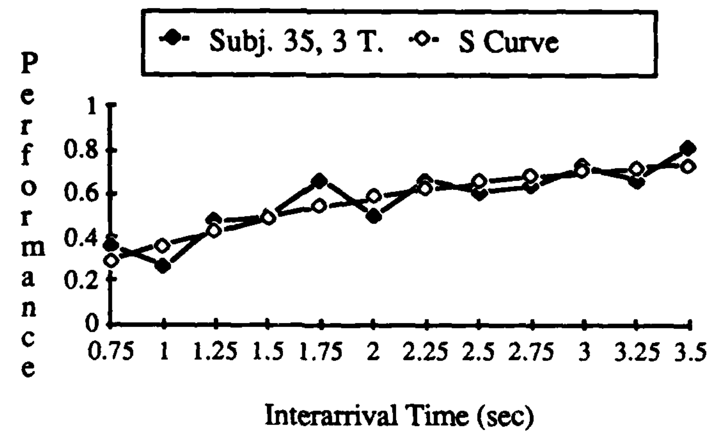
The three parameters a, b, and c of the Gompertz curve were derived for each subject for trials of both three and six tasks. The parameter 'a' ranged from 0.42 to 0.83, the parameter 'b' ranged from 1.61 to 222.78, and 'c' ranged from 0.77 t o 7.15. The distribution of the values for parameter 'b' was *not* uniform: for trials of three tasks, 23 of the 'b' values were less than 25.00 whereas for trials of six tasks, there were 22 'b' values less than 25.00. The large values taken by the parameter 'b' for some of the subjects was due to the following reasons. First, performance J is not very sensitive to changes in b. Second, a very small convergence criterion was used in the regression. Finally, by combining equations 5.2 and 5.3, b may be expressed as the logarithm of the ratio of the performance at T equal zero, to the performance as T tends to infinity. Therefore, if the subject's performance for very small T values is 0 or close to 0, b will be very large. In the five cases when the parameter 'b' was exceptionally large, for the lowest T values, the subjetcs' performance was very close to 0.

In every case the Gompertz fit was good: the min R^2 was 0.93, and a check of the residuals showed no consistent pattern which could indicate that the Gompertz was not an appropriate model. Also, in every case, the Gompertz fit was at least as good and almost always significantly better than a straight line fit: R^2 ranged from 0.93 to 0.99 for the growth curve, whereas for the linear regression, R^2 varied from 0.45 to 0.93. A one sided statistical t test was made to verify that the R^2 for the Gompertz fit were significantly larger than that for the linear fit. The t value obtained was 23.7. It is much larger than the maximum t* value which would confirm that the two distributions are not significantly different. ($t^*_{0.95,24}$=1.078 for a one sided test with a 0.95 level of confidence and 24 degrees of freedom.). In section 5.4.1, the characteristics of the data were described as being similar to the characteristics of the Gompertz curves. These observations, combined with the large R^2 values for every subject indicate that the Gompertz curves are a good description of the data. The t test confirms the Gompertz' good fit as well as the existence of a time threshold T* (which will be evaluated in section 5.4.2): The bounded rationality constraint exists.

Figure 5.4 show the Gompertz fit superimposed on the observed data. The subjects and the number of ratios are the same than the ones used for Figure 5.3.



(a)



(b)

Figure 5.4 The Gompertzt Fit for Two Subjects

5.4.3 Evaluation of T*

The existence of F_{max} was proved for every subject. Before testing the stability of F_{max}, procedures to evaluate T* are needed. This section describes how T* may be found both analytically and graphically.

In order to stay as close as possible to the theoretical model, (the corner point of the 'single step' function), T* was defined as the point at the intersection of the following tangent lines: the asymptotic performance (the parameter 'a' of the Gompertz curve), and the slope at the inflection point of the Gompertz curve. (See Figure 5.5). The first line forces performance to be at maximum, whereas the other is a good approximation of the speed at which the subject reaches maximum performance as T increases. Had the slope between the maximum and minimum asymptotes been constant, that slope would have been chosen. Figure 5.5 shows the tangent lines and resulting T* value for the same S curve as shown in Figure 5.4 a.

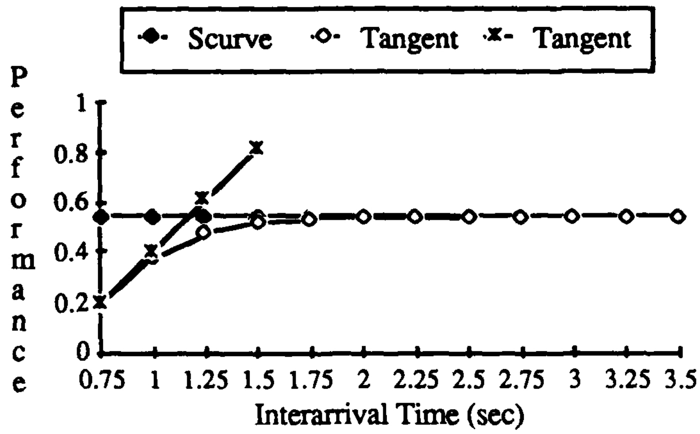


Fig.5.5 Construction of T* using Tangents

Analytically, T* may be also found as the intersection of the two lines:

$$\begin{cases} J = a \\ J = \alpha\ T^{*} + \beta \end{cases} \tag{5.12}$$

where a is the asymptote of the Gompertz fit. Therefore:

$$T^* = (a - \beta) / \alpha \tag{5.13}$$

where α is the slope at inflection point and β is intercept of the tangent at the inflection point. Since, φρομ Equations 5.4 and 5.5,

$$\alpha = a\,c / e^1, \qquad J_{inflection} = a / e^1, \qquad T_{inflection} = \ln(b) / c,$$

then,

$$J_{inflection} = \alpha\, T_{inflection} + \beta \tag{5.14}$$

$$\beta = a\ (1 - \ln(b)) / e^1 \tag{5.15}$$

Substituting α and β in Equation 5.13, the folowing expression for T^* is obtained:

$$T^* = [\, e^1 - 1 + \ln(b)\,] / c \tag{5.16}$$

where b and c are two of the three parameters of the Gompertz curve.

It is interesting to notice that the asymptote of the Gompertz curve, the parameter a, is not present in the equation. The sensitivity of T^* with respect to a is nonetheless larger than that with respect to b or c, since a is related to T^* through b and c by the Gompertz model. Further computations have shown, as expected, that T^* is more sensitive to a than it is to b or c.

5.4.4 The Stability of F_{max} Across Similar Tasks: T^*_3 versus T^*_6

For each subject i, T^*_i was computed for both three and six tasks and noted $T_i{}^*{}_3$ and $T_i{}^*{}_6$. The obtained T^* values are summarized in Table 5.3. Both the mean value and the standard deviations were very similar for three and six tasks: 2.079 sec. versus 2.069 sec. for the mean and 0.651 sec. versus 0.579 sec. for the standard deviation .

Table 5.3 Summary of T^* Values (sec.) for Three and Six Tasks

	Mean	Std. dev	Min.	Max.
Three Tasks	2.079	0.651	0.911	4.046
Six Tasks	2.069	0.579	1.080	3.504

Generally, the subjects had T^* values for three and six tasks that were very close. A little over half of the subjects, thirteen out of the twenty-five, had a larger T^* for three tasks than for six tasks. Also, since the mean of T^* over subjects were very close for three and six tasks -- only a 0.01 difference -- one was tempted to conclude that there was no systematic difference in the T^*'s as a function of the number of ratios. To confirm such a hypothesis, a few tests had to be performed. First, one had to check that the two distributions were of the same type, and then, that their mean was not significantly different.

The slightly larger standard deviation of the $T_3{}^*$ distribution was mostly due to one significantly larger $T_3{}^*$ value: 4.046 sec. The subject who had a high $T_3{}^*$ was not performing especially worse for three than for six tasks but the performance was increasing more irregularly. He had complained about the setting of the experiment, and reported entering several times the wrong answer because of inadvertently pressing the wrong key.

A plot of the distribution of the T^*'s for three tasks (Figure 5.6) and for six tasks (Figure 5.7) leads to the hypothesis that the two distributions are normal. It is interesting to note that in the case of three tasks, most of the difference between the T^* distribution and the normal distribution is due to the fact that the distribution of the T^*'s is extremely peaked. In the case of six tasks, the difference is caused both by the smaller T^* values as well as by the peak around the mean. The Chi-Square test consists of evaluating the difference (noted Q^2) between the distribution under study and (in this case), the normal distribution; Q^2 is computed as follows:

$$Q^2 = \sum_{i=1}^{5} (\text{Observed}_i - \text{Expected}_i)^2 / \text{Expected}_i \tag{5.17}$$

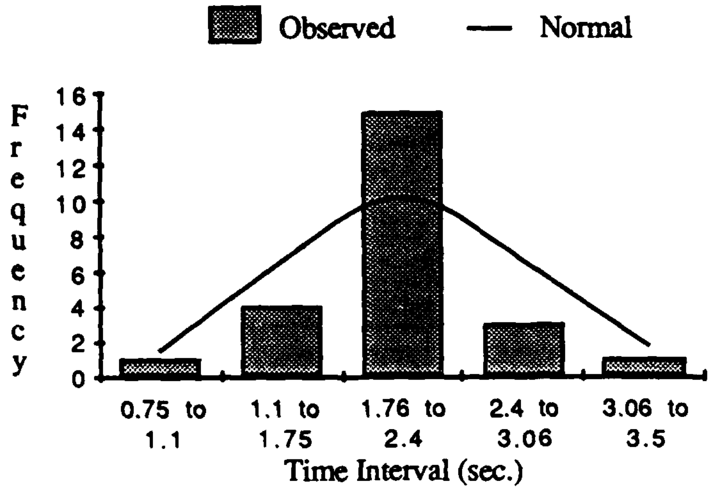


Figure 5.6 Distribution of the T* Values for Three Tasks

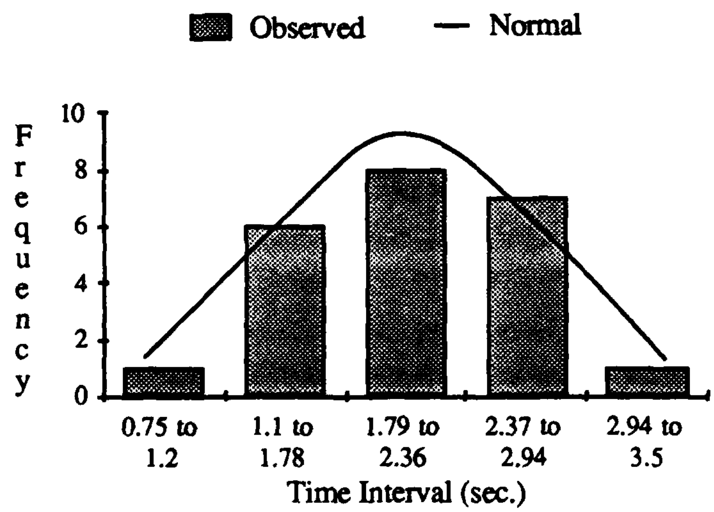


Figure 5.7 Distribution of the T* Values for Six Tasks

The Q^2 values were 5.6 for three tasks and 4.4 for six tasks which were both smaller than the critical value: $\chi^2_{,0.95,3} = 5.99$. Thus, it could be concluded that the two distributions were both not significantly different from a normal distribution, and were of the same type.

The next step was to compare the mean value of the T* distribution for three and for six tasks. A statistical test, the t test, was run. (The test performed is the t test used when comparing two dependent samples.) The t value obtained was 0.09 ($t = 0.09 < t^*_{23,.95} = 1.74$) which confirms the hypothesis that the two distributions were not significantly different.

Therefore, it may be concluded that T^* is robust with respect to minor task changes, and assuming that the workload for six tasks is approximately twice that for three tasks, the same may be postulated for F_{max}. As a result, each subject i was assigned a single value T_i^* which was equal to the average of $T_i^*{}_3$ and $T_i^*{}_6$. The frequency distribution of the individual T_i^*'s was plotted. (See Figure 5.8). This distribution is unimodal, very peaked, and has mean 2.074 sec. and standard deviation 0.549 sec.

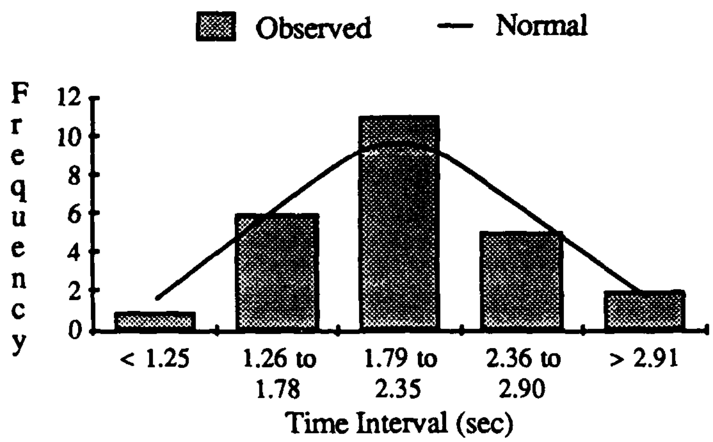


Fig. 5.8 Distribution of the Average T_i^* Values.

The distribution of the T_i^*'s for three and that for six tasks was shown to be normal. Such was also the case for the individual T^* values: A χ^2 test for goodness of fit revealed non-significant deviation from normality:

$$Q^2 = 4.4 < \chi^2(.95,2) = 5.99.$$

The fact that the T^* distribution is normally distributed is of interest since one may postulate that F_{max} for each subject will also be normally distributed. If this postulation is confirmed in Chapter 8 by the analytical results, then the hypothesis that F_{max} is stable across subjects will be validated.

5.5 CONCLUSIONS

The existence of the bounded rationality constraint, F_{max}, has been proved by the experimental results. T*, the time threshold associated with the bounded rationality constraint, has been evaluated for each subject and both numbers of tasks. It was shown that the T* value for three and six tasks were not significantly different. Therefore, under the assumption that the workload for six tasks is approximately twice that for three tasks, one may conclude that F_{max} is stable when minor task changes are made. Finally, a T* value was estimated for each subject. The distribution of the individual T*'s was normal. Such a result enables the postulation that F_{max} is stable across subjects.

The stability of F_{max} both across similar tasks, and across subjects will be confirmed in Chapter 8 when both the experimental and analytical results are combined. First, however, models of the algorithms used by the subjects are presented in Chapter 6. Then, in Chapter 7, the workload associated with these algorithms is evaluated.

6. THE DECISIONMAKING MODEL: THE SUBJECTS' VIEWPOINT

6.1 INTRODUCTION

The goal of this project was to study the bounded rationality constraint F_{max}. Such a study requires both experimental and analytical results. In Chapter 5, the experimental results were described: the existence of F_{max} was proved, T* was evaluated for each subject, and statements were made about the stability of F_{max} across tasks. The next goal of this report is to present the analytical results, (the computation of workload), and confirm the assumptions raised in Chapter 5 concerning the stability of F_{max}. To compute the workload associated with the task, the subjects' mental process must be modeled and then transformed into information-theoretic algorithms. This chapter presents basic mathematical models of the subjects' mental process.

A mathematical model attempting to describe the subjects' mental process would be of little significance if it was not validated. Therefore, it seemed appropriate to evaluate the appropriateness of these models. After running the experiment, the subjects were asked to describe the algorithm(s) that they had used while running the experiment; these results are described in section 6.2. The major dififculties encountered when modeling the tasks are described in section 6.3. Then, simple mathematical models which took into account the algorithms described by the subjects were developed and are presented in section 6.4. Each subject was assigned to a particular algorithm. Before analyzing these models and computing the workload associated with each (Chapter 7), the appropriateness of the algorithms is evaluated in section 6.5. The performance of the models is compared to that of the subjects.

6.2 SUBJECTS' STATEMENTS

6.2.1 Correspondence with Cognitive Science

From reading the subjects' description of the algorithms used, as well as their general comments about the experiment, it appeared that the subjects felt under time pressure, and that they had been using coping strategies to perform the task. The task was to compare ratios and find which was the smallest. To ensure 100% performance, a computer program would have processed the task by computing the value of each ratio and then comparing the obtained values. It appeared that the subjects often only processed a portion of the input information that they would normally use, if they had more time or aids (even pen and pencil) to perform the task. Subjects used shortcuts and filtering methods that allowed them to processes the most significant information. Examples of such behavior were subjects who systematically ignored the second digit of the two-digit values of speed and distance. Such an observation is similar to the conclusions drawn from the few studies of time pressure found in behavioral decision literature (Wright, 1974).

6.2.2 Retrieving Descriptions of the Model(s) Used

As it was mentioned in the previous section, the subjects were asked to describe the algorithm that they had used to perform the task. Before the subjects' statements were studied, different models that would be plausible descriptions of the algorithms were designed. These models were used as guidelines when the descriptions were too vague.

The first task was to translate the subjects' description into a mathematical model. Whereas some subjects seemed able to analyze very clearly the basic mental processes that they have used, others seemed unable to do so. Phrases like 'When the comparison is not obvious...' appeared more often than expected. A study of the rest of the description often gave some idea of the algorithm (or at least the algorithmic structure) used. Here are a few extracts of some of the subjects' answers:

Extract A:

Step 1: Observe left hand column of multi digit fractions

Step 2: Try to look for 8's or 9's in the second column

Step 3: When digits on the left are the same, decide based on second column digits

Extract B:

For ratios <1 compare numerators if the ratios comparable, otherwise obvious

For ratios >1 if comparable try and reduce otherwise want smaller numerator, greater denominator.

The models were aggregated into a few categories which are discussed in section 6.4. Translating the subjects' description required a subjective methodology where both intuition and 'common sense' played a very important role. Such modeling methods required an evaluation of each algorithm using some test of appropriateness or some other evaluation method. Such a test, which was alluded to in the first section of this chapter, is described in detail, in section 6.5.

6.2.3 The Stages of the Decision Process

In Chapter 2, the decision-making model was described as a two stage process. The first stage, the Situation Assessment stage, allowed the decisionmaker to analyze and assess the situation before making a decision in the response selection stage. At each stage, the subject could choose from a set of algorithms to process the information.

When running the experiment, the subjects seemed to be using only one situation assessment algorithm. The algorithm consisted of looking at the clock and understanding how much time they had to compare the ratios, understanding how many ratios would have to be processed, and finally just looking at the value of the ratios present on the screen. The subjects did not mention these first steps which are the obvious steps that one would follow when faced with such a task.

The response selection algortithm varied from subject to subject. It appeared, however, that most subjects used the same algorithm, whatever the input ratios were. The main factor which seemed to induce a change in algorithms was the time allotted to perform the task. When they could not process the task using the strategy they were most comfortable with or their 'optimum strategy', subjects often switched either to a simpler version of the same algorithmic structure, or to a different structure. The subjects were instructed not to guess unless it was an educated guess, but subjects sometimes just picked one of the two ratios randomly, often hoping that the next comparison would be easier. Changes in strategies due to increase in time pressure were very difficult to monitor since most subjects were not even aware of the change, or if they were, did not report it.

As a result, the models that were derived for each subject, encompass both the Situation Assessment and the Response Selection Stages, but do not take into account the subjects' relationship with the clock. For this particular experiment, the two stage decision model of the single decisionmaker shown in Figure 2.6 may be simplified as in shown in Figure 6.1.

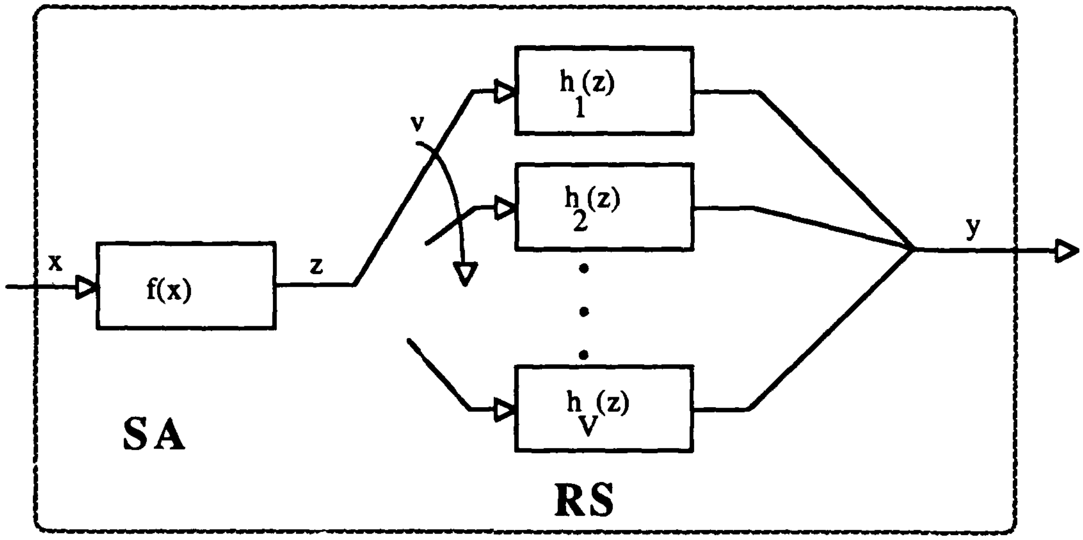


Figure 6.1. The Simplified Decision-Making Model

6.2.4 The Issues of Pure and Mixed Strategies

In the case of this experiment, when considering the type of strategies used by the subjects, the notions of pure and mixed strategies as described in Boettcher and Levis (1982) seem difficult to apply. Pure and mixed strategies are defined as follows. In the case of the situation assessment stage, a decisionmaker without a preprocessor uses a pure strategy if whatever the input, he uses a given algorithm to process that input with probability one, (he always uses the same situation assessment algorithm). In the case of the RS stage, the notion is very similar. For each input identified by the situation assessment stage, there is only one response selection algorithm that the DM will use to provide a response. This may be expressed mathematically as follows:

$$p(v = j \mid z = z_i) = 1 \tag{6.1}$$

where j is the algorithm selected in the response selection
z_i is the output of the situation assessment algorithm.

In the experiment, it was very difficult to evaluate which strategy or algorithm(s) the subjects were using. It was even more so when trying to identify which subject changed algorithm when. Because of the experimental setup, as explained in the previous section, (6.2.3), there was only one situation assessment algorithm, thereby there could only be a pure strategy. For the response selection stage, the setup did not force the subjects to use any particular algorithm. From talking to the subjects and reading their comments, it appeared that the subjects used a single strategy whatever the input was. It is only when they felt too *pressured* that they switched from their 'usual' strategy to a simpler one. Therefore, since the change of strategies was based on one of the input characteristics, (the time available to process the trial), they were using a set of pure strategies for the response selection stage.

6.3 MODELING DIFFICULTIES

6.3.1 Requirements of Information Theory

As described in Chapter 2, information theory is a mathematical tool which may be used to compute the cognitive workload associated with a given task. Information theory imposes constraints and requirements on the type of tasks that may be modeled as well as on the algorithms that may be used. These conditions restrict the type of tasks that may be simulated.

One of the major constraints is that the tasks be well defined so that they can be modeled using mathematical variables, or at least variables for which a probability distribution may be derived. As a result, the quantities and parameters which are used must be measurable values, and belong to a finite set.

The other conditions which must be fulfilled are that the decisionmakers be well trained and motivated and that they operate at a level where the bounded rationality is not in effect. The last condition concerning the bounded rationality constraint is particularly important to this section of the research and has serious implications when considering the algorithms that will be modeled to compute the cognitive workload. It has been metioned that subjects have been switching from one algorithm to an other as the time allotted per trial was decreased. When subjects felt overloaded, or close to being overloaded, many switched to an algorithm for which the cognitive workload was less; these algorithms were called coping algorithms. As a result, when modeling the task and assessing the workload, it will be very important to model the algorithm that subjects used when they did not feel under serious pressure yet, i.e., the algorithm that they used when they have the most time available.

The growth curves which were used to model the experimental data smoothed out any change in strategy. Therefore, T* may be considered as an average over several 'T*', each 'T*' associated with an algorithm requiring less cognitive workload: a coping strategy. Since the individual 'T*'s were not identifiable, the T* value (see Equation (5.16)) was retained. It may also be postulated, that the slope at which performance decreases, (more specifically the slope at the inflection point), reflects the number of different coping algorithms used by the subject as the time available to perform the task decreased: the larger the number of different algorithms used, the smaller the slope, and consequently, the smaller the T* value.

6.3.2 The Limitation of the Mathematical Models

Information theory restricts the type of algorithms that may be used as well as the experimental setups. One of the major problems in trying to assess the mental workload is also derived from the difficulty or better the incapacity to include non-quantitative measures in the mathematical models. How may one model a subject's mental process when the subject describes choosing one ratio over another because 'the comparison was obvious', or how can one describe the fact that another subject will just assume that 2/5 is less than 3/7 ? In both cases, the subject knows (or thinks he knows) the answer and uses some cognitive process to make a decision. No previous research has been done to evaluate and compute using information theory the cognitive workload associated with intuition. The impact of memory on workload has been discussed in the literature (Hall, 1982; Bejjani, 1985). In this research, for simplicity, it is assumed that the decisionmakers are memoryless with respect to short term memory. Also, with respect to long term memory, the only cognitive work which is assessed when choosing the smallest of two single digit ratios is due to the distribution of each ratio. The cognitive work required to retrieve the information from permanent memory is ignored but could be the subject of future research.

6.4 THE RESULTING MODELS

6.4.1 The Different Mental Approaches

When considering all the constraints imposed by the analytical tools as well as by the nature of the task, the number of different approaches was quite small. It appeared that there were only three different basic types of mental processes. Whereas some features were common to all three types, the most important processing in each case was quite different. The three different methods were the following:

Method 1. For each ratio, approximate the speed and distance with single digit values, then compare the resulting ratio.

Method 2. Approximate the ratio (or its inverse) to its nearest integer and compare.

Method 3. Compare the differences between numerators and denominators.

Whereas for the first two methods the first steps could be done independently for each ratio, the last approach included both ratios as soon as some processing was done. Each method resulted in one, two or three different algorithms to include some of the variability among subjects. The resulting set of models consisted of six different algorithms that will be described in detail in the next section. Finally, before performing any computation or approximation, it appeared that the subjects checked for any significantly small ratio. If such a ratio was spotted, they ignored the other ratios and would give the 'small ratio' as the solution. Such a procedure was even more widely spread when the time allotted per comparison was small. For small processing times, the notion of a small ratio was often less strict, and included ratios that would not have been considered if the clock had shown more time available.

6.4.2 The Six Algorithms: Description of the Models.

Models derived from method 1

The first approach (method 1 described above), which consisted of approximating the last digit of both speed and distance, was used by four subjects. Two different algorithms resulted from this approach. The first approximation method, (named Algorithm 1), was to simply truncate the last digit of both speed and distance values when performing the comparison. The second method, (named Algorithm 2), is to truncate first the last digit of the speed and distance values as for Algorithm 1, and then add to the truncated values 0 if the second digit is less than 5 and 1 if the second digit is larger than 5. Once the ratio values have been approximated, the subject has to compare the two resulting ratios. If the two are not equal, the solution is the smallest ratio. If the two are equal, the subject randomly picks one of the two as a solution. Given two input ratios R1 and R2 such that

$$R1 = d1/v1 \quad \text{and} \quad R2 = d2/v2,$$

one comparison for Algorithm 1 is described in Figure 6.2, whereas one comparison for Algorithm 2 is described in Figure 6.3.

d[1]=trunc[d1/10]
d[2]=trunc[d2/10]
v[1]=trunc[v1/10]
v[2]=trunc[v2/10]

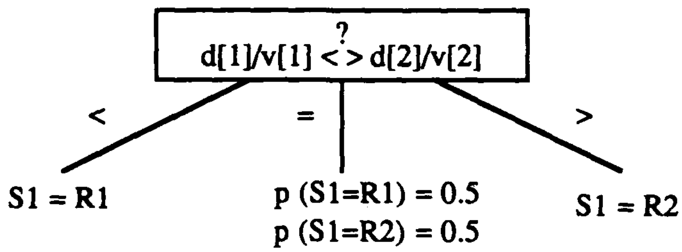


Figure 6.2 One Comparison Using Algorithm 1

d[1]=round[d1/10]
v[1]=round[v1/10]

d[2]=round[d2/10]
v[2]=round[v2/10]

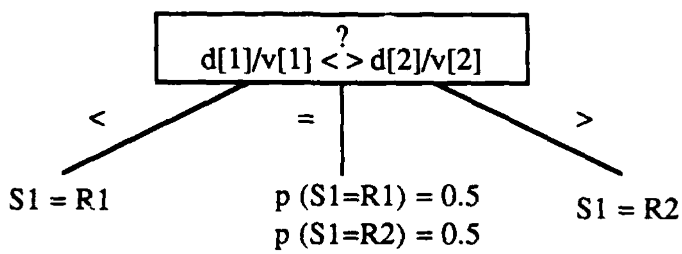


Figure 6.3 One Comparison Using Algorithm 2

Models derived from method 2

Only one algorithm was derived from method 2. This model had the disadvantage of being different for ratios that were less than one and for ratios that were larger than one. For ratios that were larger than one, each ratio was rounded to its nearest integer. Then, if the absolute difference between the nearest integer and the ratio was more than 0.25, the integer value was corrected by positive 0.25 or by negative 0.25, as appropriate. Then the resulting values for both ratios were compared. As for algorithms 1 and 2, if the values were the same, it was assumed that the subjects picked randomly one of the two ratios for the solution. For ratios less than one, the inverse of the ratio is first taken. Then, the same process as for ratios larger than one is used. The resulting algorithm was called Algorithm 3 and the process for one comparison is shown in Figure 5.4 for ratios larger than one and in Figure 5.5 for ratios less than one. Considering the two ratios R1 and R2 already defined for algorithm 1 and 2, Algorithm 3 is described for ratios larger than one in Figure 6. 4 and for ratios less than one in Figure 6.5.

Ratios > 1 $\qquad$ i = 1, 2

rat[i] = round[d_i/v_i]
if [rat[i] - (d_i / v_i)] > 0.25 $\qquad$ then ratio[i] = rat[i] - 0.25
if [rat[i] + (d_i / v_i)] > 0.25 $\qquad$ then ratio[i] = rat[i] + 0.25

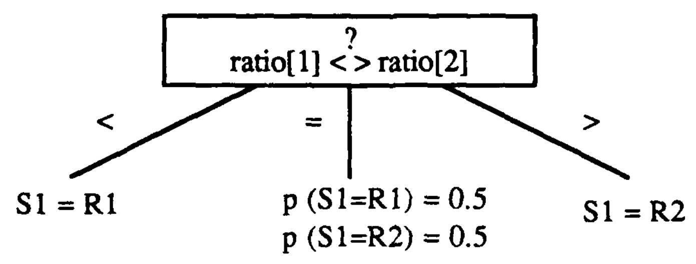


Figure 6.4 One Comparison Using Algorithm 3 for Ratios Larger than One

Models derived from method 3

Three algorithms were derived from method 3 which consisted of comparing the differences between the numerators and denominators (distances and speeds) of the two ratios that had to be compared.

Ratios < 1 $\quad i = 1, 2$

$rat[i] = round[v_i/d_i]$

if $(1/rat[i]) - (d_i / v_i) > 0.25$ then $ratio[i] = rat[i] - 0.25$

if $(1/rat[i]) + (d_i / v_i) > 0.25$ then $ratio[i] = rat[i] + 0.25$

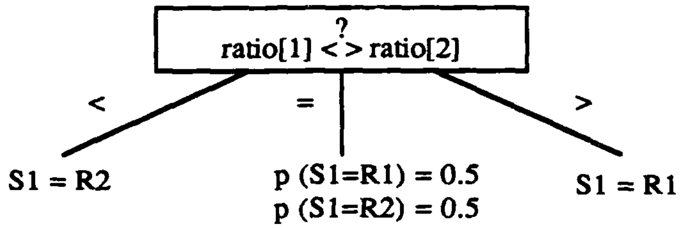


Figure 6.5 One Comparison Using Algorithm 3 for Ratios Less than One

For Algorithm 4 and Algorithm 5, the difference between the distance and the speed of each ratio was computed, then, the ratio with the smallest difference was chosen. For Algorithm 4, the subject could come to a conclusion if the difference was larger than 10. For Algorithm 5, the subject came to a conclusion if the difference between the speeds was larger than that between the distances or vice versa. The two algorithms are described below in Figures 6.6 and 6.7.

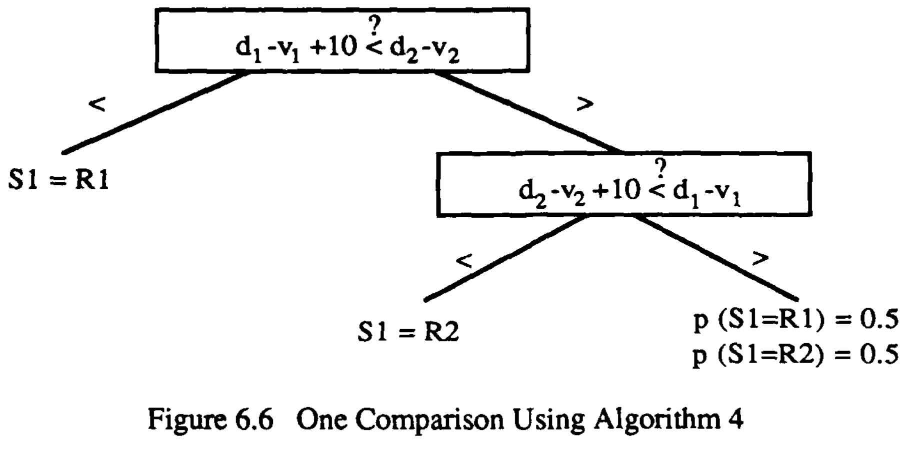


Figure 6.6 One Comparison Using Algorithm 4

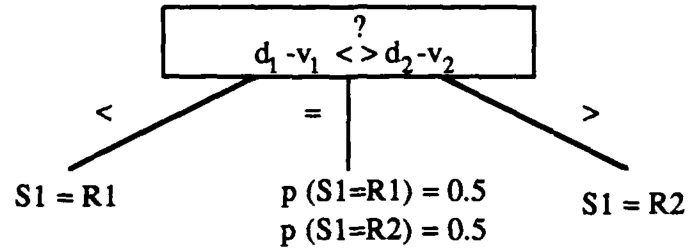


Figure 6.7 One Comparison Using Algorithm 5

The last model, Algorithm 6 is a combination of Algorithm 2 and method 3. The subject first checks if there is not one ratio which has a smaller distance and a larger speed than the other. If he can not make a decision by these criteria, the subject uses the approximation method of Algorithm 2. Algorithm 6 is described in Figure 6.8.

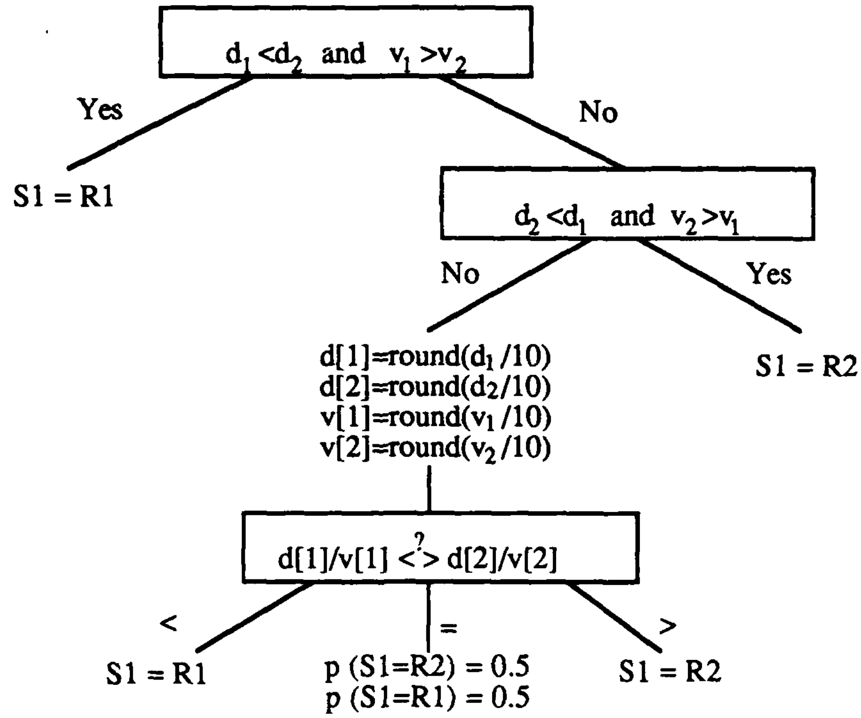


Figure 6.8 One Comparison Using Algorithm 6

6.5 EVALUATING THE MODELS

6.5.1 Purpose of the Evaluation

The different models used by the subjects have just been described. However, before assuming that these models are a reasonable representation of the subjects' mental processes, the appropriateness of these models must be validated. To do so, the maximum performance of each subject will be compared to the estimated performance of the algorithm associated with each subject.

6.5.2 Defining the Maximum Performance

Each subject's maximum performance was established from the experimental results using the S curves. For subject i, the maximum performance is noted a_{i3} for three tasks and a_{i6} for six tasks, and may be derived as follows:

$$\text{for } j = 3 \text{ and } 6 \qquad a_{ij} = \lim_{T \to \infty} \left(J_{ij}(T) \right) \tag{6.2}$$

Each of the six algorithms described in section 6.4 represents a pure strategy and is noted f_k, with k taking values ranging from 1 to 6. For a given algorithm f_k, the estimated performance will be noted J_{k3} for three tasks and J_{k6} for six tasks.

The performance that would result from accurately using these algorithms has been estimated by simulating the experiment 300 times on an IBM PC. Each algorithm was programmed in Pascal, and the function "random" was used to generate sets of ratios satisfying the requirements of the experiment, the same way the experiment had been set up. But since whether the sets of ratios were less or larger than one depended on another random function, it seemed important to simulate the experiment for both ratios (larger than and less than one) for the same number of times.

Such a procedure gave particularly relevant information concerning the difficulty of the experiment. Some subjects had mentioned that they found the ratios larger than one more dificult to compare than the ratios less than one. This observation was confirmed by the simulation of the algorithms: the algorithms always performed significantly better for the ratios less than one.

Since the trials were independent identically distributed Bernoulli variables, the estimated performance J_{kj} could be computed as follows:

$$\bar{J}_{kj} = (\bar{J}_{kj<1} + \bar{J}_{kj>1}) / 2 \tag{6.3}$$

where:

$$\bar{J}_{kj<1} = \sum_{i=1}^{150} x_{i<1} \; / \; 150 \qquad J_{kj>1} = \sum_{i=1}^{150} x_{i>1} \; / \; 150 \tag{6.4}$$

However, since each subject's performance curve had been transformed using the arcsine transformation to perform the regression analysis, it was necessary to make the same transformation on the algorithms' expected performance to have values that could be compared. Therefore, an arcsine transformation was made on the algorithms' simulated performance. Table 6.1 shows the (transformed) ***estimated*** performance for each of the algorithms for three tasks and the non transformed performance both for ratios less than one and ratios larger than one. Table 6.2 shows the same results for six tasks.

The results are only estimates of the population's true mean. The variance for each estimated performance was very low. It varied between 0.0005 to 0.005. (The sample size was 300 of a population of possible combinations of ratios close to 10^{13}.)

The algorithms' estimated performance values were larger for trials of ratios less than one than for trials of ratios larger than one. The difference may be explained by the constraints imposed on the trials. For trials of ratios less than one, the values of the ratios were constrained so that the difference between any two ratios be at least 0.05. The same constraint was not imposed on trials of ratios larger than one for practical reasons: when running trials of six tasks, the program often could not generate ratios satisfying the constraints. Instead, the ratios larger than one were constrained to be larger than 1.2. As a result, the ratios larger than one were on average slightly harder than the ones less than one.

Table 6.1 Estimated Performance for the Six Algorithms for Three Trials

Algorithm number	Estimated Performance (Three Trials) Ratios < 1 untransf.	Ratios >1 untransf.	Overall Perf. untransf.	Overall Perf. arcsine transf.
Al.1	0.84	0.625	0.733	0.654
Al.2	0.86	0.645	0.753	0.665
Al.3	0.91	0.724	0.817	0.719
Al.4	0.744	0.437	0.591	0.558
Al.5	0.757	0.628	0.693	0.627
Al.6	0.86	0.705	0.783	0.692

Table 6.2 Estimated Performance for the Six Algorithms for Six Trials

Algorithm number	Estimated Performance (Six Trials) Ratios < 1 untransf.	Ratios >1 untransf.	Overall Perf. untransf.	Overall Perf. arcsine transf.
Al.1	0.645	0.538	0.592	0.559
Al.2	0.657	0.584	0.621	0.580
Al.3	0.774	0.427	0.601	0.564
Al.4	0.608	0.349	0.479	0.486
Al.5	0.632	0.462	0.547	0.530
Al.6	0.832	0.591	0.711	0.639

Figure 6.9 shows the estimated performance of each algorithm for both three and six tasks. The algorithms perform better for three than for six tasks, but the ordering of the algorithms' performance stays almost unchanged. (Algorithm 3 which performed the best for three tasks, is only third to best for six tasks. The others have remained unchanged). The average difference between performance for three tasks and performance for six tasks is a 0.1 decrease. Finally, Figure 6.9 also shows that the difference in performance among the algorithms is not very large.

For three tasks, there is only a 0.16 difference between the best and the worst algorithm, the difference is 0.15 for six tasks. However, considering the small variances of the algorithms' estimated performance (in the range of 10^{-3}), the differences should not be considered as negligible.

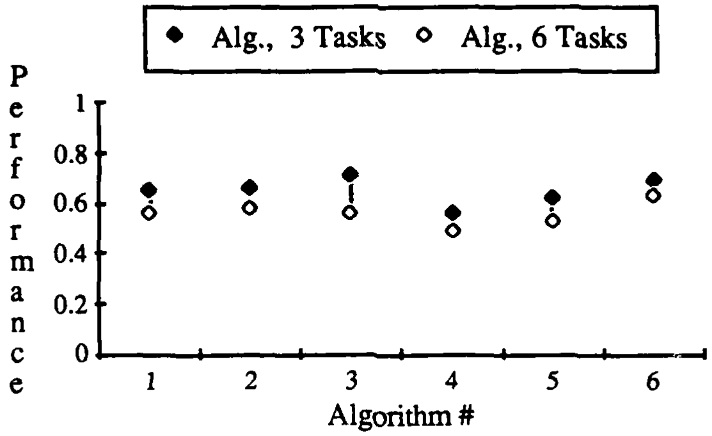


Figure 6.9 Algorithm Performances: Three Tasks versus Six Tasks

6.5.3 Comparing Performance: Simulations versus the Experiments

The six algorithms described in section 6.4 were derived from the subjects' descriptions. Each subject was then assigned to the algorithm which was closer to the description he gave. The next step was to estimate the algorithms' maximum performance. The goal of this section is to evaluate the appropriatness of the algorithms.

Table 6.3 shows, for three tasks, the number of subjects who were using each algorithm, the average performance over the subjects and, finally, the algorithm's performance (The subject's performance which was averaged was the asymptotic performance, the 'a' values of the Gompertz fit, see Equation (6.2)). Table 6.4 shows the results for six tasks. The difference between the algorithms' and the subjects' performance was within a close range for three tasks; this is shown explicitly in Figure 6.10.

Three subjects performed significantly better than the algorithms that they seemed to have been using. These subjects were in the School of Engineering and had had very high scores on the SAT's and the GRE's. They seemed very familiar with approximation methods, therefore one may hypothesize that when the algorithms they were using could not give a significant conlusion, they made educated guesses.

Table 6.3 Three Tasks: Subject Performance Versus Algorithm Performance

Algorithm #	Number of Subjects Using it	Average Perf. Over the Subjects	Algorithm's Estimated Perf.
1	2	0.573	0.654
2	3	0.590	0.665
3	6	0.715	0.719
4	3	0.555	0.558
5	4	0.655	0.627
6	7	0.682	0.692

For six tasks, Table 6.4 suggests that, on average, the subjects were performing better than the algorithms which were modeled. Since not a single subject mentioned using a different algorithm for three than for six tasks, the algorithms were considered to be satisfactory models.

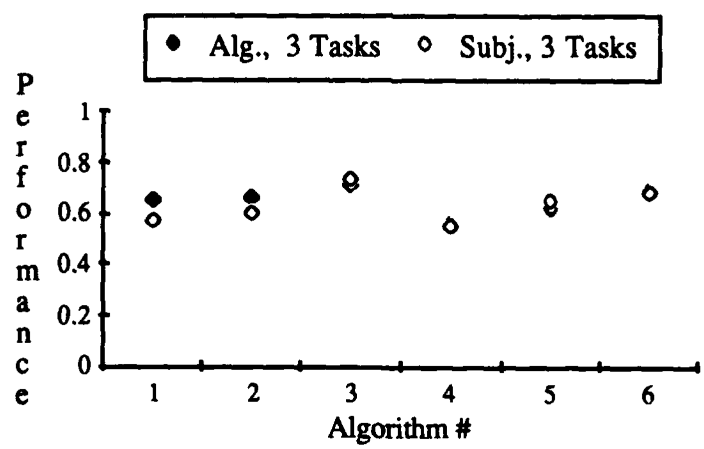


Figure 6.10 Subject Performance Versus Algorithm Performance: Three Tasks

Table 6.4 Six Tasks: Subject Performance Versus Algorithm Performance

Algorithm #	Number of Subjects Using it	Average Perf. Over the Subjects	Algorithm's Estimated Perf.
1	2	0.543	0.559
2	3	0.688	0.580
3	6	0.732	0.564
4	3	0.585	0.486
5	4	0.645	0.530
6	7	0.704	0.639

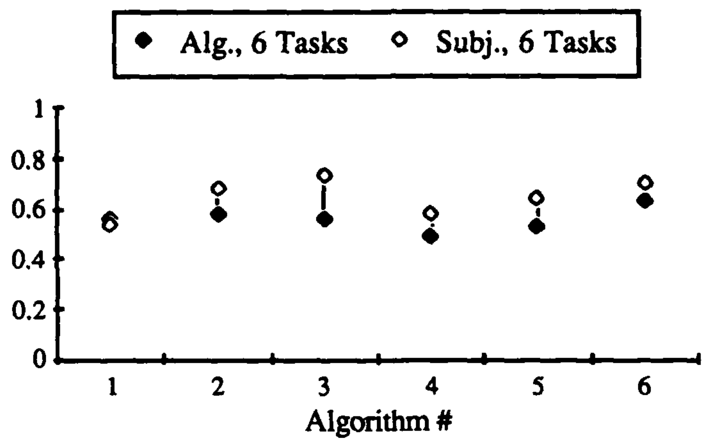


Figure 6.11 Subject Performance Versus Algorithm Performance: Six Tasks

Overall, the obtained results were satisfactory; the next step is to compute the workload associated with each algorithm and estimate F_{max} for each subject.

7. WORKLOAD EVALUATION

The workload for the different algorithms is evaluated in this chapter. The information theoretic model of each algorithm is obtained and the entropy of each variable is computed. Thus, the workload associated with each algorithm can be evaluated.

The different steps of the modeling process are described in section 7.1. First, the input alphabet is characterized, but it is impossible to enumerate. Then, the internal variables are reviewed. In particular, the level of detail needed, and the effects of temporary and permanent memory on the assessment of workload are studied. Finally, the impact of having trials of ratios either larger than one or less than one is discussed. Section 7.2 describes the steps followed to compute the entropy of the different variables. Finally , the workload is evaluated in section 7.3. First numerical values of the workload of the different algorithm are given, then the feasibility of these values are discussed and the experimental and analytical results are compared.

7.1 THE INFORMATION-THEORETIC ALGORITHMS

7.1.1 The Input Alphabet

The input alphabet is first defined for both numbers of ratios. Then the size of the alphabets and the input entropies are estimated.

When the subjects start the experiment, the following information is available to them on the computer screen: the number of ratios that are to be processed for the trial, the amount of time they will have to process the task and, finally, the distance and the speed of the two ratios that they will first have to compare. (See Figure 4.4). The time available to perform the task is a parameter which varies from trial to trial.

It is assumed that the amount of cognitive workload required both to acknowledge the amount of time available to perform the task and to register the time available is negligible compared to the workload necessary to process the tasks. Therefore, the input vector includes only the information about the number of ratios and the value of the speeds and distances of these ratios. As a result, the input vector to trials of three tasks consists of a set of four ratios, whereas the input vector to trials of six tasks consists of a set of seven ratios. Each threat is actually a pair of speed and distance values. In case of three tasks, such an input vector noted

x_3 will be described as follows:

$$x_3 = (d1/v1, d2/v2, d3/v3, d4/v4) \tag{7.1}$$

where d1, d2, d3, and d4 are the distances associated with ratios 1, 2, 3 and 4, and v1,v2, v3 and v4 are the speeds associated with the same ratios. An example of such an input vector may be the following:

$$x_{3,i} = (11/ 34, 25/89, 32/33, 28/57) \tag{7.2}$$

The values taken by the distances and the speeds are constrained by the requirements described in section 4.5. There are three types of sets: First, the set S_1 of possible speeds and distances, then R_1, the set of possible ratios where the speeds and distances belong to S_1. Finally X_3, and X_6, are the sets of possible combinations of ratios for three and six tasks. X_3 and X_6, may also be divided into subsets of ratios larger than one and subsets of ratios smaller than one, noted $X_{3|x<1}$, $X_{3|x>1}$, $X_{6|x<1}$, $X_{6|x>1}$, respectively.

The input alphabets are X_3 for trials of three tasks and X_6 for trials of six tasks. The ordering of the components of each input vector matters, i.e., the two vectors $x_{3,1}$ and $x_{3,2}$ are not considered identical.

$$x_{3,1} = (11/ 34, 25/89, 32/33, 28/57) \tag{7.3}$$

$$x_{3,2} = (25/89, 11/34, 32/33, 28/57) \tag{7.4}$$

The above vectors are different because the order in which the subjects process the ratios often has an impact on the final solution. The subjects use approximation methods to compare the ratios; as a result, when given the same ratios but in a different order, the probability of error is affected.

The input alphabets have been characterized. Now the distribution and the number of elements of the input alphabets X_3 and X_6 must be evaluated to compute the entropy of the input vectors x_3 and x_6.

The distribution of both alphabets is assumed to be uniform, because each input vector x_i is generated randomly. (It is assumed that each vector has the same probability of being generated.). The cardinal of each input alphabet is difficult to assess because of the constraints imposed on the ratios. Therefore these figures are estimated as follows. First the number of elements of each alphabet is computed assuming that there are no constraints on the sets of ratios. Then, a computer program is used to estimate the number of eligible combinations of ratios when the constraints are included.

The pool of acceptable ratios less than one is 3003, and the pool of acceptable ratios larger than one is 2407. (These figures were computed by generating every possible pair of distances and speeds and counting all the feasible ones. The number of ratios larger than one is less than the number of ratios less than one, because the ratios larger than one were subject to an additional constraint: they had to be larger than 1.2).

If the constraints imposed among combinations of ratios were ignored, the number of input vectors less than one for three tasks would be:

$$A^{4}_{3003} = 3003 * 3002 * 3001 * 3000 = 8.1162 * 10^{13} \qquad (7.5)$$

and the number of input vectors larger than one would be:

$$A^{4}_{2407} = 2407 * 2406 * 2405 * 2404 = 3.3483 * 10^{13} \qquad (7.6)$$

That is, ignoring the constraints imposed betweeen ratios, the size of the input alphabet X_4 would be:

$$A^{4}_{3003} + A^{4}_{2407} = 11.4645 * 10^{13} \qquad (7.7)$$

The same way, ignoring the constraints imposed betweeen ratios, the size of the input alphabet X_6 would be:

$$A^{7}_{3003} + A^{7}_{2407} = 2.187 * 10^{24} + 4.634 * 10^{23} = 2.650 * 10^{24} \qquad (7.8)$$

Such large input alphabets do not allow enumeration.

The program which was used to estimate the number of feasible input ratios was based on the method used to generate sets of ratios during the experiment. An iteration consisted of picking a distance and a speed satisfying the necesary constraints. Then the number of possible second ratios was computed by enumeration. A second ratio out of the pool of possible ratios was then picked randomly, and the number of possible third ratios was then computed... Following the same procedure for the remaining ratios, for each run, the program computed the number of possible second N_2, third N_3, fourth $N_4..N_7$ ratios. For each run i, for three tasks the number of possible combinations of ratios, noted P_{i3} could be derived as the following product:

$$P_{i3} = N_{i1}*N_{i2}*N_{i3}*N_{i4} \tag{7.9}$$

and for six tasks P_{i6}:

$$P_{i6} = N_{i1}*N_{i2}*N_{i3}*N_{i4}\, N_{i5}*N_{i6}*N_{i7} \tag{7.10}$$

The program was run 150 times for both ratios larger than one and ratios less than one. The estimated number of of possible first, second, third ..seventh ratios were derived for ratios larger than one and for ratios less than one for both number of ratios as follows:

Ratios <1

$$\bar{N}_{j\,|<1} = \left(\sum_{i=1}^{150} N_{ij\,|<1} \right) / 150 \qquad j = 1 \text{ to } 7 \tag{7.11}$$

Ratios >1

$$\bar{N}_{j\,|>1} = \left(\sum_{i=1}^{150} N_{ij\,|>1} \right) / 150 \qquad j = 1 \text{ to } 7 \tag{7.12}$$

Therefore, the size of the input alphabet X_3 could be derived as following:

$$C_{x3} = \prod_{i=1}^{4} \bar{N}_{i4\,|<1} + \prod_{i=1}^{4} \bar{N}_{i4\,|>1} \tag{7.13}$$

The results for three tasks were the following:

$$C_{X3} = (3003*2567*2163*1793) + (2407*2355*2315*2276) \quad (7.14)$$

$$C_{X3} = 2.9896*10^{13} + 2.9867*10^{13} = 5.9763*10^{13} \quad (7.15)$$

The size of the input alphabet X_6, noted C_{X6} was derived using the same method as for X_3. The results were as follows:

$$C_{X6} = (3003*2567*2163*1793*1459*1161*913)$$

$$+ (2407*2355*2315*2276*2238*2202*2168) \quad (7.16)$$

$$C_{X6} = 4.6236*10^{22} + 3.1910*10^{23} = 3.6534*10^{23} \quad (7.17)$$

The constraints imposed on the set of ratios also created difficulties when considering the internal variables which are described in section 7.1.2.

7.1.2 The Internal Variables

Before considering the entropy of the internal variables and the workload associated with each algorithm, the internal variables must be characterized. Therefore, as a first step, the subjects' approach to the experimental task and the level of detail used for modeling the algorithms are defined. Then the methodology used to assess the probability distributions of the internal variables is described.

Two different approaches were possible when modeling the experiment The subjects' tasks could be interpreted either as : 'to find the smallest ratio of a population sample' or as 'given four ratios, find the smallest'. In the first case, the distribution of the value of the smallest ratio when observing samples of four would have been the critical issue. In the latter case, the values of the smallest ratio would have been of no importance. Instead, the smallest ratio's position in the sequence (that is what is the first, second, third or fourth) would have been the required solution. The first approach was modeled in this thesis. The stategies that the subjects reported using were influenced by the values the ratios could take. Therefore models based on population samples seemed more appropriate. Another modeling issue related to short term and long term memory. With regard to short term memory, it is assumed that the decisionmakers are memoryless: they do not remember the approximated value of the ratio which was smaller in the previous comparison and must approximate it again for the following comparison. Such an assumption was derived after talking to subjects. They reported that they generally reestimated the ratios for each comparison. With regard to long term memory, it was assumed that the subjects could rank order the single digits ratios and did not need any special algorithm to do so.

The modeling approach has been discussed and the level of detail used in the models is now described. Within each algorihm, the different processes are kept as steps, but each operation required to perform the process is not recorded as a variable. This methodology keeps the number of internal variables under control; only the basic variables are recorded as variables. The internal variables of the first decision of Algorithm 1 for three tasks are described below in Figure 7.1, as an example.

The notation used in Figure 7.1 may be described as follows:

dij = jth digit of distance of ratio i. dij ranges from 1 to 9

$$w21 = \min(Ti,Tj) = \begin{cases} 0 \text{ if the two values are the same} \\ 1 \text{ if the first is smallest, Ti in this case} \\ 2 \text{ if the second is smallest, Tj in this case} \end{cases}$$

w22 = distance associated with w21, where w22 takes the value of the distance associated with the ratio corresponding to the value of w21.

If w21 had taken a value of 1, w22 would take the values of d(Ri), since Ri would be smaller than Rj; such a ratio could be noted Ri'. If w21 takes a value of 0, each ratio (either Ri or Rj) has a probability of 0.5 of being chosen.

The modeling process and the choice of internal variables have been described. The next step is to derive the probability distribution of each variable and compute the workload of each algorithm. First, however, the impact of two of the experimental setups on the probability distributions are discussed. The effect of having trials consisting of ratios either larger than one or less than one is described in section 7.1.3. Then, the assumptions required to evaluate the probability distributions are described in sections 7.2 and 7.3.

Input vector, X   X=(d1/v1, d2/v2, d3/v3, d4/v4)

Internal Variables, wi

w1 = d1	w5 = v1	w9 = trunc(d1/10) = d11	w13 = trunc(v1/10) = v11
w2 = d2	w6 = v2	w10 = trunc(d2/10) = d21	w14 = trunc(v2/10) = v21
w3 = d3	w7 = v3	w11 = trunc(d3/10) = d31	w15 = trunc(v3/10) = v31
w4 = d4	w8 = v4	w12 = trunc(d4/10) = d41	w16 = trunc(v4/10) = v41

IF   w17 d1<20 and v1>90   THEN   Y = R1 END OF ALGORITHM
ELSE IF   w18 d2<20 and v2>90   THEN   Y = R2 END OF ALGORITHM

ELSE   w19 = d11/v11 = T1
          w20 = d21/v21 = T2

          w21 = min(T1,T2)
          w22 = distance of w21 = d(w21)
          w23 = speed of w21 = v(w21)

NEXT COMPARISON

Figure 7.1 The Information Theoretic Description of Algorithm 1: The First Decision

7.1.3 The Trials: Ratios Less than One and Ratios Larger than One

The trials were set up so that whether the ratios would be larger than one or less than one would be picked randomly. Such a setup had an impact on the distribution of the internal variables. There was a 0.5 probability that a trial would consist of ratios less than one, and a 0.5 probability that the trial would consist of ratios larger than one. Therefore, the entropy of an internal variable wi may be expressed as follows:

$$H(wi) = -\sum_{wi} p_{wi}(wi) \log_2 p_{wi}(wi) \qquad (7.18)$$

where

$$p_{wi}(wi) = p_{wi|x<1}(wi|x<1)*p(x<1) + p_{wi|x>1}(wi|x>1)*p(x>1) \qquad (7.19)$$

x is the ratio from which w_i is derived

$$p(x<1) = p(x>1) = 0.5 \qquad (7.20)$$

If a variable wi can *only* be derived *either* from a ratio larger than one, or from a ratio less than one then exactly one of the two equations below holds (7.21 or 7.22).

$$p_{wi|x<1}(wi|x<1) = 0 \tag{7.21}$$

or

$$p_{wi|x>1}(wi|x>1) = 0 \tag{7.22}$$

The input vector, X, as well as the individual ratios (di / vi) are such variables. For such variables Equation (7.18) may be rewritten as:

$$H(wi) = -\sum_{wi|x<1} p_{wi}(wi) \log_2 p_{wi}(wi) - \sum_{wi|x>1} p_{wi}(wi) \log_2 p_{wi}(wi) \tag{7.23}$$

Finally, Equation (7.23) for the input entropy or the entropy of the ratios may be simplified as follows:

$$\begin{aligned} H(wi) = &-\sum_{wi|x<1} p_{wi|x<1}(wi|x<1) * p(x<1) \log_2 [p_{wi|x<1}(wi|x<1) * p(x<1)] \\ &-\sum_{wi|x>1} p_{wi|x>1}(wi|x>1) * p(x>1) \log_2 [p_{wi|x>1}(wi|x>1) * p(x>1)] \end{aligned} \tag{7.24}$$

As a result, the input entropy for three tasks becomes:

$$H(x) = 0.5 * \log_2 (2.9896*10^{13}) + 0.5* \log_2 (2.9867*10^{13}) + 1 \tag{7.25}$$

$$H(x) = 22.3825 + 22.3818+1 = 45.764 \text{ bits} \tag{7.26}$$

Because of the experimental setup, for each variable, the distribution must be derived seperately for the input vectors of elements larger than one and those of elements less than one: two different probability distributions are obtained. Then, the two are combined as in Equation (7.19) to evaluate the entropy of each variable of the algorithms..

7.2 THE COMPUTATION OF ENTROPY

7.2.1 The Approach

The internal variables have been described and some of the computational issues were raised in the previous section. This section describes the methodology followed to assess the entropy of each variable.

A normal procedure to compute the probability distribution of each internal variable is to use a computer program simulating a binning process to assess the histogram of each internal variable as all the possible inputs are fed to the program. The probability distribution is then derived from the histogram. For this particular case however, a binning process using every element of the input alphabet may not be used because of the size of the input alphabet. Therefore assumptions must be made to estimate the probability distribution of each variable. First the two "categories" of internal variables are described. Then, the methodology to estimate the probability distribution is reviewed for each.

7.2.2 The Different Types of Variables

Two different types of variables may be identified within each algorithm: The variables for which the entropy may be computed without comparing two ratios, and the variables for which the entropy could only be computed after one or more of the comparisons were made. For simplicity, the first group will be called the static variables and the second the non-static variables. (In Figure 7.1, variables w1 to w18 are considered as static, whereas variables w19 to 23 are non-static.)

The static variables are variables that are repeated, and are the same for each four (or seven) ratios. The distribution of the static variables were computed for one ratio, taking all the possible ratios larger and less than one. Then the same distribution was assumed for each ratio. These variables reflect the size of the input, and as a result dominate when considering the entropy of the total system. The very large entropy of these variables tends to overshadow the decision variables of the algorithms.

The non-static variables describe three categories of variables: the decision process, the approximated value of the ratios which were chosen to be the smallest after a comparison, and the intermediate variables used to arrive at the approximated value. The probability distribution of each category of non-static variables was estimated using computer programs. The distribution of the non-static variables changes after each comparison.

7.2.3 The Entropy of the Static Variables: Assumptions and Methodology

In this section, the most important assumptions used to compute the entropy of the static variables are given, while the methodology used to compute the entropy of a few static variables is described.

The first static variables to be considered are the ratios before they are compared. The distribution among ratios less than one is assumed to be uniform. The same is valid for the ratios larger than one. This assumptions is used even though the constraints imposed on the ratios will make some ratios appear in sets more often than others. Let $\mathbf{R}$ be the pool of all feasible ratios, $\mathbf{R}_0$ the pool of all feasible ratios less than one and $\mathbf{R}_1$ be the pool of all feasible ratios larger than one. Then the above assumptions may be described as follows:

$$\forall\ r \in \mathbf{R},\ \ p\,(r \in \mathbf{R}_0) = 0.5 = p\,(r \in \mathbf{R}_1) \tag{7.29}$$

$$\forall\ r_a \in \mathbf{R}_i,\ \forall\ r_b \in \mathbf{R}_i,\ p_r(r_a) = p_r(r_b) \quad \text{for } i = 0, 1 \tag{7.30)}$$

Also, the entropy associated with each ratio of a set x = (R1, R2, R3, R4) is assumed to be the same. It is assumed that the entropy of the ratios is independent from the order the ratios appear on the screen. The entropy for each ratio may be computed as follows:

$$H_R = -\sum_{R} p_R(R) \log_2 [\, p_R(R)] \tag{7.31}$$

where $R \in \mathbf{R}$

$$H_R = 0.5 \log_2 (3003) + 0.5 \log_2 (2407) + 1 = 12.39 \text{ bits} \tag{7.32}$$

The distances and the speeds forming each ratio are the next static variables studied. It is assumed that the distances are independent from one another, but are not independent of the speed associated with them to form a ratio. The probability distribution among the different possible distance values is not uniform. The entropy of the distances and the speeds may be computed as follows:

$$H_{wi} = - \sum_{wi} p_{wi}(wi) \log_2 p_{wi}(wi) \qquad (7.33)$$

where pwi(wi) was computed by iteration using the binning process, considering first all the possible ratios larger than one, then all the possible ratios less than one. Each time the value wi appeared, the frequency of wi was increased by one. The entropy was the following:

$$H_{wi} = 6.41 \text{ bits} \qquad (7.34)$$

where wi is a speed or a distance associated with a ratio before this ratio has been compared to another ratio.

The same procedure was done to estimate the probability distributions of the first digit of both speeds and distances.

$$H_{wi} = 3.16 \text{ bits} \qquad (7.35)$$

where wi is the first digit of a speed or a distance associated with a ratio before the ratio was compared.

It is assumed that all the internal variables derived from the speeds and distances were independent of the sequence of the ratios. (The first digits are an example of such derived internal vaiables.) Therefore, these variables are assumed to be equally distributed for all four ratios when considering trials of four ratios, and all seven ratios when considering trials of seven. For example, when considering Algorithm 1, which is shown in Figure 6.1, the sets of variables shown in Table 7.1 are equally distributed.

The probability distribution of the other static variables were derived using the binning process and the assumptions just described.

Table 7.1 Sets of Equally Distributed Variables

Variables	Corresponding Internal Variables
d1, d2, d3, d4	w1 to w4
v1, v2, v3, v4	w5 to w8
d11, d21, d31, d41	w9 to w12
v11, v21, v31, v41	w13 to w16
decide if di < 20 and vi > 90	w17, w18, w24, w32
di1/vi1, i =1 to 4	w19, w20, w28, w36

7.2.4 The Entropy of the Non-Static or Decision Variables: Methodology

The distribution of the non-static variables was computed differently for each algorithm, since these variables were algorithm-specific. However, the same terminology may be used to describe the steps that were followed.

Within each algorithm, the first two ratios noted $R_1$and R_2 were approximated into T_1 and T_2 which are the variables compared for the first decision, D1. It is assumed that T_1 and T_2 are equally distributed. The distribution of the decision D1, as well as that of the minimum of T_1 and T_2 was found by first assessing the distributions of T_1 and T_2, then, finding the probability that T_1 would be smaller and finally by finding the probability distribution of the minimum of T_1 and T_2. The same procedure was continued until the fourth or seventh approximated ratio was compared to the minimum of the previous comparison. While such a procedure was followed to find the distribution of the decision variables, the same method was used to assess the distribution of the 'non-static' variables.

The probability that the approximated ratio x1with distribution p_{x1} be less than the approxiamted ratio x2 with distribution p_{x2} was computed as follows:

$$p(\, x1 < x2) = \sum_{\text{all } x1} p_{x1}(x1) \sum_{x1}^{\infty} p_{x2}(x) \qquad (7.38)$$

The distribution of the min of two variables x1, x2, was computed as follows:

$$y = \min(x1, x2) \tag{7.39}$$

$$P_y(y) = p_{x1}(y) \sum_{y}^{\infty} p_{x2}(x2) + p_{x2}(y) \sum_{y}^{\infty} p_{x1}(x1) \tag{7.40}$$

These formulas were used to compute the entropy of the non-static variables of the different algorithms.

7.3 THE WORKLOAD FOR EACH ALGORITHM

This section first summarizes the most important assumptions regarding the assessment of the variables' probability distribution. Secondly, the numerical values of the workload are presented and discussed. Thirdly, the feasibility of the results is reviewed by checking the consistency between the algorithms. Finally, the assumption derived in Chapter 4 regarding the correspondence between the workload for three and for six tasks is discussed. The evaluation of workload allows the testing of the hypotheses concerning the bounded rationality constraint which is presented in Chapter 8.

7.3.1 The Most Important Assumptions

Many assumptions and approximations have been described in section 7.2. Each has been used in the computation of the total entropy of the appropriate algorithm(s) to evaluate the workload associated with each algorithm. The most important and the most critical were the following:

(1) Assume uniform distribution of the input.

(2) Assume uniform distribution of the ratios, i.e., each ratio has the same probability of occurring in an input.

(3) The distribution of the approximated ratios and all the intermediate steps to obtain the approximated ratios is based on the first two assumptions.

(4) After a given comparison, the rate of change in entropy of the similar types of non-static variables is assumed to be the same. The rate of change is defined as the ratio of the entropy of the non-static variable used for comparison i to the entropy of the same variable when used for comparison i-1. (Examples of

similar types of non-static variables would be the first digits and second digits of the speed values, or the actual distance values and the approximation of the distance values used to make the comparison.)

7.3.2 The Numerical Values

The workload for each number of ratios and each algorithm was computed following the methodologies described in section 7.2. The numerical values are summarized in Table 7.2. As one may see from the table, the value of the workload varies significantly from algorithm to algorithm. For three tasks the workload ranges from 165.62 bits to 275.58 bits and the mean is 235.03. For six tasks, it ranges from 297.92 to 513. 59 bits and the mean is 433.04 bits.

Table 7.2 The Workload Associated with the Algorithms

Algorithm	Workload Three Tasks (in bits)	Workload Six Tasks (in bits)
1	210.103	386.700
2	262.031	480.059
3	275.582	513.594
4	227.858	417.450
5	165.615	297.915
6	268.995	502.530

The variation among algorithms is weighted by the number of subjects who were associated with the algorithm. In Chapter 6, each subject was assigned an algorithm which attempted to model the basic operations or approximations performed by the subject. Therefore, the average (over the subjects) workload required by the experiment may be computed by multiplying the number of subjects who "used" a given algorithm by the workload of this algorithm. The results, when considering the number of subjects associated with each algorithm, are summarized in Table 7.3.

Table 7.3 The Average Workload for the Experiment Over Subjects

	Three Tasks	Six Tasks
Average workload	243.625	450.270
Standard Deviation	40.353	79.057

7.3.3 Consistency Among the Algorithms

When looking at the workload for both three tasks and six tasks, the workload associated with Algorithm 5 is signficantly lower than that of the other algorithms (165.615 bits for three tasks and 297.915 bits for six tasks). Such a low workload is explained by the structure of the algorithm itself. The algorithm consists of comparing the difference between the speeds and distances of the two ratios. Such a process requires only two steps before making the comparison i.e., compute each difference, which drastically reduces the workload. The workload is not based on the number of steps, but on the entropy associated with each variable. Because many of the intermediate internal variables have very significant entropies, the number of intermediate steps required to transform the input into variables that may be compared plays a significant role in the total entropy. Such an observation is particularily true for Algorithm 5, which is very simple. It is also applicable to Algorithm 1 which requires a limited number of steps before the comparisons are made.

Algorithm 1 has a larger workload than Algorithm 5 (210.103 bits for three tasks, and 386.700 bits for six tasks versus 165.615 and 297.915 bits) but it is still lower than that of the other three algorithms. Six steps are required to transform two input ratios into two variables that may be compared: truncate each speed and each distance (4 steps), and then form each single digit ratio (two extra steps). The other algorithms require a significant number of steps before a comparison is made.

The fact that Algorithms 1 and 5 have smaller workload than the other three is explained by their structure. Another method to check the results of the workload values is by looking at the three different categories of algorithms which were derived in Chapter 6.

The first category included Algorithms 1 and 2 in which the ratios were transformed into single digit ratios and were compared. Algorithm 2 was defined as requiring more processing than Algorithm 1 since for the first case the rounded ratios are compared whereas in the other case the truncated ratios are compared. The computations of workload confirmed the expectations, the workload for Algorithm 2 is larger than that for Algorithm 1 (210.103 bits versus 262.031 bits for three tasks and 386.700 bits versus 480.059 bits for six tasks, an increase of 24.7 % for three tasks and 24.1 % for six tasks).

The second category of algorithms included algorithms 4, 5 and 6. The workload for Algorithm 4 is larger than that for Algorithm 5. The same structure is used, but Algorithm 4 computes four differences as opposed to two and makes two comparisons as opposed to one. The increase of workload was very significant, 37.6% for three tasks, and 40.1% for six tasks. Such an increase could be expected since the amount of internal processing is almost doubled. Algorithm 6 is a combination of Algorithms 2 and 5. It uses the first steps of Algorithm 5 to determine if a small ratio could be spotted before any computation. If the test is not relevant, it rounds each ratio using the same methodolgy as Algorithm 2. The workload for Algorithm 6 was slightly larger than that for Algorithm 2 as expected, (268.995 bits versus 262.031 bits for three tasks, and 502.530 bits versus 480.059 for six tasks.) The increase of 2.8% for three and 4.6% for six tasks is small. The testing variables used in Algorithm 6 (and not present in Algorithm 2) have entropies of a few bits only.

Finally, Algorithm 3 is a seperate category since a different strategy is used for ratios less than one and larger than one. As a result, the number of internal variables is significantly increased even though each comparison requires only six intermediate variables (as Algorithm 1), two of which have entropies less than 2. Because of the different strategies for ratios less and larger than one, the workload for Agorithm 3 is the largest of all.

From the above remarks, it appears that the values for the workload are consistent between the algorithms. As a result the relative differences between the workload of the different algorithms are feasible and conclusions relating the different algorithms and their 'users' may be derived based on these values. The next step is to compare the workload for the same strategies, but for the different number of tasks within a trial.

7.3.4 Comparing the Workload for Three and Six Tasks

In Chapter 4, it was postulated that the important parameters were not the number of ratios but the number of tasks. The assumption was: the workload per comparison is approximately the same for three and six tasks i.e., the workload for six tasks should be twice that for three tasks. The experimental results seemed to confirm this assumption since the T* values for three and six tasks were not significantly different. This section first shows the ratio of workload for three and six tasks for each algorithm. Then the values obtained are discussed and explained, and the validity of the assumption is assessed. Finally, a simple linear regression modeling the workload as a function of the number of tasks is presented .

The analytical results confirm the assumption that the workload for six tasks is approximately twice that for three tasks. On average, the ratio of the workload for six tasks to that of three tasks is close to 1.84. Table 7.4 shows the ratio for the six algorithms as well as the average over the six algorithms and the average when introducing the frequency of each algorithm.

Table7.4 The ratio of the Workload for Six Tasks to that of Three Tasks

	Algorithm #	Ratio (Six Tasks / Three Tasks)	Average Over Subjects
	1	1.841	1.845
	2	1.832	
	3	1.864	
	4	1.799	
	5	1.868	
	6	1.887	
Average Over Algorithms	1.839		

The fact the the workload for six tasks is not twice that for three tasks should not be regarded as unwanted noise. Such a 'discrepancy' is derived from the analytical models. First the entropy of the input is not proportional to the number of comparisons and does not increase linearly with the number of ratios because of the log function. The input for three tasks is 45.76 bits and for six 77.68 bits). Then, the internal variables increase this difference even more because the entropy of more than half of the internal variables reflect the entropy of the very large input alphabet. Finally, when considering the distribution of the minimum of two equally uniformely distributed variables (these were the assumptions used), it will be skewed towards the smallest values. This is particularly relevant to our experiment when considering the distribution of the min as the number of comparisons increases. The previous paragraph may be described analytically as follows:

Let X be an ordered population uniformly distributed and let N be the size of the population. Then

$$p_x(x) = \begin{cases} \dfrac{1}{N} & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases} \tag{7.41}$$

Let $y = \min(x_1, x_2)$ where x_1, x_2 are two elements of X, and f_y the distribution of y then :

$$f_y(y) = \begin{cases} \dfrac{2}{N}\left(1 - \dfrac{y}{N}\right) & \text{if } y \in X \\ 0 & \text{otherwise} \end{cases} \tag{7.42}$$

Let $z = \min(y, x_3)$, $x_3 \in X$ and g_z the distribution of z, then

$$g_z(z) = \begin{cases} \dfrac{3}{N}\left(1 - \dfrac{z}{N}\right)^2 & \text{if } z \in X \\ 0 & \text{otherwise} \end{cases} \tag{7.43}$$

The distribution of the variable $t \in X$ being the smallest of the n^{th} comparison and a variable $u \in X$ is:

$$f_t(t) = \begin{cases} \dfrac{n}{N}\left(1 - \dfrac{t}{N}\right)^{n-1} & \text{if } t \in X \\ 0 & \text{otherwise} \end{cases} \tag{7.44}$$

As an analogy to our experiment, x_1 and x_2 may be assumed to be the first two ratios to be compared. y takes the values of the ratios kept from the first comparison, x_3 is the third ratio to be compared, z, takes the values of the ratios kept from the second comparison, etc. The distributions become more and more skewed, thereby reducing the entropy of the minimum after each comparison.

The decrease in entropy after each comparison ranges between 2% and 5% of the non-static variables. This is not very significant when considering the entropy of the whole system and the entropy of the static variables which are not affected by the decrease due to the comparisons. Also, in this particular case, the entropy related to the large input tends to dominate the entropy of the system and absorb the changes due to the decrease of the entropy of the decision variables (called non-static variables).

A simple least squares fit using the twelve data points of Table 6.2 (three and six tasks, Algorithms 1 through 6),

$$Y_i = a X_i + b \tag{7.45}$$

where

$X_i = 3, ..., 3, 6, ..., 6$
$Y_i = 210.03, 262.031..., 268.995, 386.700, 480.059, ...502.530$

yields

$$Y = 66 X + 37 \tag{7.46}$$

For

$X = 3 \qquad Y = 235$
$X = 6 \qquad Y = 433$

Note that the constant 37 is equivalent to about half the effort of a comparison and is not very significant either for three or six comparisons. Because of the very few data points used (twelve), this regression should only be considered as a gross model, but it is important to note that the results are consistent with the other observations.

Therefore, considering all the assumptions which have been made throughout this thesis, the analytical results do not contradict the experimental results. The assumption made in Chapter 5 was reasonable: the workload per comparison is approximately the same for three and six tasks.

The workload was evaluated for each algorithm and the values were consistent both between algorithms and with the experimental results. Therefore, these values may be used to assess the bounded rationality constraint for each subject and test hypotheses about the stability of F_{max} both across subjects and across tasks.

8. THE BOUNDED RATIONALITY CONSTRAINT: RESULTS AND ANALYSIS

The bounded rationality constraint for each subject and its behavior are presented in this chapter. First, the hypotheses regarding the stability of F_{max} are stated. Then, the methodologies used to evaluate F_{max} and to test the hypotheses are described. Next, F_{max} is evaluated for each subject and each type of trial, i.e., for three and six tasks. Finally the validity of the hypotheses is tested and the results are compared to the statements made in Chaper 5.

8.1 THE HYPOTHESES

Two hypotheses concerningthe stability of F_{max} are to be confirmed.

Hypothesis (1). F_{max} is stable for an individual when minor tasks changes are made.

Hypothesis (2). F_{max} is stable across individuals and across tasks.

8.2 METHODOLOGIES

8.2.1 The Procedures to Evaluate F_{max}

In Chapter 5, the minimum average time required to perform the experiment was derived for each subject using the experimental results. In Chapter 7, the workload associated to each model was evaluated. The bounded rationality constraint which is noted F_{max} may now be computed for each subject and for both types of trials combining the experimental and the analytical results.

As described in section 4.1 , F_{max} is the ratio of ~~the workload~~ associated to the trial to the time threshold T*. Since the values of T* were evaluated as a time per task, the value of T* has to be multiplied either by three or six to consider the total duration of the trials. Therefore, for each subject and for both number of tasks, the value for the bounded rationality constraint may be computed as follows:

$$F_{max\ ,i,j} = G_{\ i,j} / [\ j * T^{*}_{i,j}] \tag{8.1}$$

where

i is the subject number and j is the number of tasks

$G_{\ i,j}$ is the workload of the algorithm associated to subject i for j tasks

$T^{*}_{i,j}$ is the threshold processing time associated to subject i for j tasks

8.2.2 The Procedures for Testing the Hypotheses

The methodologies used to test the hypotheses are very similar to the methodologies used to test for the stability of T* across trials and across subjects.

To test the stabity of F_{max} across trials, first the distributions of $F_{max,3}$ and $F_{max,6}$ are assessed using a statistical test (the Chi-Square test) and are then compared. If the two distributions are of the same type, then it is tested if the mean of the two distributions are significantly different using a statistical test, (the t test).

The second hypothesis: the stability of F_{max} across trials and subjects is more simple to confirm. First, an F_{max} value is estimated for each subject, (for each subject, F_{max} is the average of $F_{max,3}$ and $F_{max,6}$). Then, a Chi-Square test is used to estimate whether the F_{max} distribution is significantly different from the normal distribution or not. A non-significant difference would lead to the conclusion that F_{max} is stable both across subjects and tasks.

8.3 COMPUTATION OF F_{max}

The values of F_{max} were computed for each subject for both number of tasks and are shown in Table 8.1 and were summarized in Table 8.2. The average value of $F_{max\,j}$ over subjects is 44.35 bits/sec for three trials versus 41.00 bits/sec. for six trials. The standard deviation for three tasks is quite large 15, as is the one for six tasks, 13. It is interesting to note that in both cases the standard deviation is almost one third of the mean.

Table 8.1 The F_{max} Values for Each Subject and Both Numbers of Tasks

Subject #	$F_{max,3}$	$F_{max,6}$
20	42.776	30.714
21	47.036	32.636
22	83.378	64.422
23	64.838	46.516
25	25.896	23.631
26	38.380	26.350
27	45.704	43.714
28	49.510	41.220
29	28.214	22.549
31	42.719	26.839
33	31.605	29.064
34	36.016	61.100
35	27.241	35.911
36	38.124	34.798
37	30.595	31.217
38	17.310	24.954
39	44.786	44.392
41	54.397	62.652
44	65.718	55.087
45	42.096	29.775
46	28.737	23.903
50	45.150	44.840
51	31.113	42.148
52	64.684	54.414
53	40.672	42.081

Table 8.2 Summary of the F_{max} Values for Both Numbers of Tasks

	$F_{max,3}$ (bits/sec)	$F_{max,6}$ (bits/sec)
Average	42.668	38.997
St. Dev.	15.068	12.873
Min	17.310	22.549
Max	83.378	64.422

It is important to realize however, that the values obtained for the bounded rationality constraint are not of any specific interest if just considered as values. The different algorithms that could be used to model the same task could increase the workload, and therefore F_{max} as well by a factor of two or more. Therefore, it is by studying the distribution of F_{max} as the tasks is slightly changed, and across subjects, as well as by comparing the conclusions derived analytically with the conclusions derived experimentally that the significant conclusions may be derived. As long as each algorithm is modeled consistently with the others, the comparisons may be done.

8.4 TESTING THE HYPOTHESES

8.4.1 The stability of F_{max} Across Trials

To test the stability of F_{max} across trials, the distribution of $F_{max,3}$ and $F_{max,6}$ must first be evaluated. In Chapter IV, it was established that the T* values were normally distributed for both three and six tasks and it had been postulated that the distribution of the T*'s should be closely related to that of F_{max}. This postulation was confirmed: goodness of fit tests showed that the distribution of both $F_{max,3}$ and $F_{max,6}$ were normal. (The Q^2 error was 2.0 for three trials and only 0.8 for six trials). Figure 8.1 shows the distribution of $F_{max,3}$ over subjects, and Figure 8.2 shows the frequency distribution of $F_{max,6}$. The difference between the normal distribution and that of the $F_{max,3}$ values is shown in Figure 8.1, whereas the difference between the normal distribution and the $F_{max,6}$ values is shown in Figure 8.2. The size of the intervals is not the same. The intervals are constructed as for the Chi-Square test: the cumulative probability within each interval is 0.2.

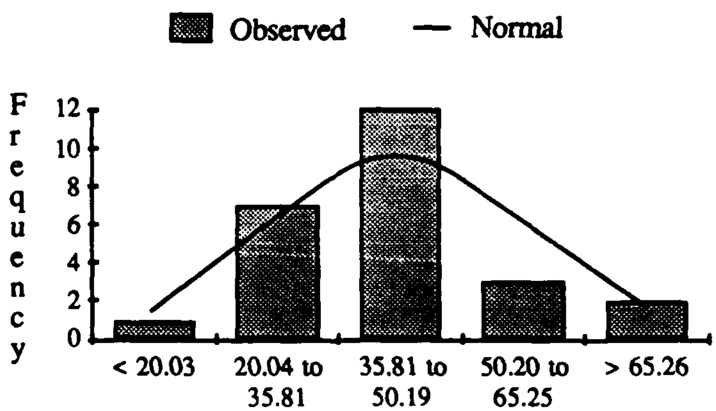


Figure 8.1 The Distribution of F_{max} for Three Trials

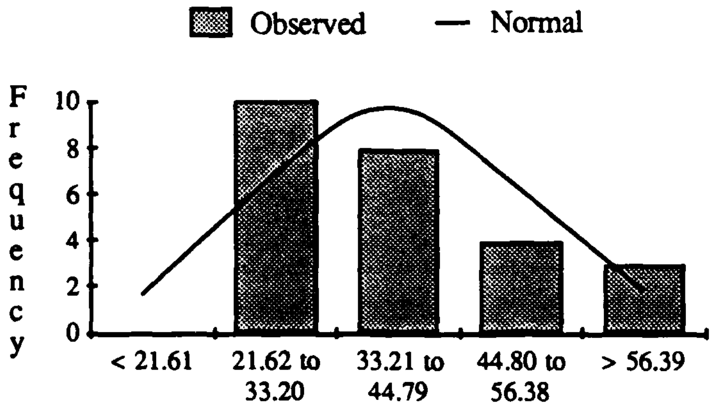


Figure 8.2 The Distribution of F_{max} for Six Trials

The next step needed to validate the hypothesis that F_{max} is stable across tasks is to compare the means of the $F_{max,3}$ and $F_{max,6}$ distributions. The experimental results had postulated that F_{max} was not significantly different for trials of three and six tasks. This result was confirmed by a statistical t test. The value for the statistical t test was 1.79 . The critical value for a two sided t test at a 0.95 level of confidence with 24 degrees of freedom is 2.06; 2.06 is larger than 1.79, so the hypothesis that the two distributions are of same mean may not be refuted .

Therefore, one may say that F_{max} is stable for each subject as the number of tasks is varied from three to six. As a result, it may be assumed that there is only one significant value for each subject, which will be taken as the average of the F_{max} 's for three and six tasks.

In addition, these results provide indirect evidence for the stability of F_{max} over time, since each subject was tested on three or four different days. (A "composite" curve resulting from wide day to day fluctuations in the bounded rationality constraint would not likely reveal a clear threshold.) This stability suggests that it may not be necessary to measure a decision maker's F_{max} value for every type of task the decision maker may have to perform. Instead, the decision maker's F_{max} value could be measured using a prototypic "calibration" task. The value obtained from this prototypic task could be safely assumed to apply to a substantial range of structurally similar tasks.

8.4.2 The Stability of F_{max} Across Subjects

The next step of this Chapter is to study the behavior of F_{max} over all subjects. The F_{max} associated with each subject i was computed as follows:

$$F_{max,i} = \sum_{j=3,6} F_{max,i,j} \,/\, 2 \qquad \text{for } i = 1 \text{ to } 25 \qquad (8.2)$$

The F_{max} values were summarized in Table 8.3. A Goodness of fit test showed that the distribution was not significantly different from normal (the error is $Q^2 = 5.2 < \chi_{0.95,2} = 5.99$). Therefore, it may be assumed that the distribution of F_{max} over subjects is stable, and the analytical results confirm the experimental results. Figure 8.3 shows the disribution of the individual values of F_{max}.

The analytical results have confirmed the experimental results. The bounded rationality not only exists for all the subjects, but it is uniformly distributed for each type of trials over the subjects, it is stable to minor tasks changes, and finally it is also uniformly distributed when assuming only one F_{max} value for each subject.

Table 8.3 Summary of the Average F_{max} Values over Subjects (in bits per sec.)

Mean	40.830
Standard Deviation	13.013
Min	21.132
Max	73.906

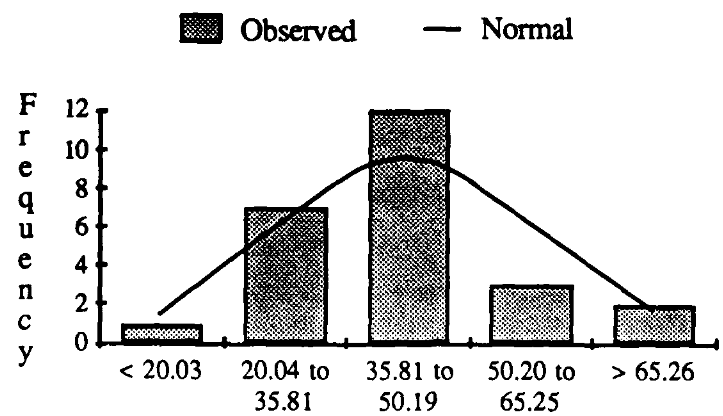


Figure 8.3 Distribution of the Average F_{max} Values over Subjects

When considering a particular task performed by well trained decisionmakers, it may be assumed that despite the individual differences and the different algorithms used, the bounded rationality is uniformly distributed among people. One could submit the hypothesis that in a very strict environment such as the military, where people who perform the same job should all be very qualified, the distribution of individual bounded rationality constraint for similar tasks would not only be normal but also extremely peaked. This could help significantly when designing organizations where the decisonmakers are not to be overloaded.

9. CONCLUSIONS AND FUTURE RESEARCH

Both the analytical and experimental results were needed to answer most of the questions related to the bounded rationality constraint of human decision makers. The first significant results are derived from the experimental analysis in Chapter 5. The existence of a threshold beyond which performance degrades rapidly, namely, the bounded rationality constraint, was established. Then, on the basis of the first result, two hypotheses were formulated. One dealt with the stability of the bounded rationality constraint across similar tasks and the other with the stability across subjects. The experimental results were combined with the results from the mathematical models to derive the value of the maximum processing rate for each subject for trials of three and six tasks, respectively. The hypotheses were then tested: the bounded rationality constraint was shown to be both stable across similar tasks and across subjects.

Information theory was the mathematical tool used to assess the amount of cognitive workload required to perform the experiment, given the different algorithms that were modeled. The workload associated with the different algorithms was consistent with the complexity of the algorithms and the different categories of algorithms. Such a result gave some validation of the mathematical model used. When trying to model the difference between the number of ratios, there was a slight discrepancy between the experimental and analytical results. Three explanations were offered to account the slight difference. First, the model for three and six tasks might not have captured the different approach that the subjects might have taken during the experiment. When assessing models in Chapter 6, it was found that simulations of the models for six tasks consistently predicted worse performance than the subjects', whereas the performance was very similar when considering three tasks. Second, considering the very large size of the input alphabet, it is possible that the subjects did not recognize that the probability distribution of some of the variables was changing as the number of ratios to consider increased; the subjects might not have changed their strategy accordingly. Third, it should not be forgotten that the experimental results, particularly the threshold values for the time intervals, were artificially constructed from the data, and therefore necessarily introduced some error. Finally, other factors such as time allocation, or short term memory may have affected the workload, but consideration of these factors was beyond the scope of this first experimental project. Because not a single subject mentioned using a different approach when processing trials of three and six comparisons, the models described in this report are reasonable considering the small discrepancy.

The existence of a bounded rationality constraint for each subject was proved from the experimental results. Performance was fairly stable before it degraded rapidly. The Gompertz (S-shaped) curves which were used to model the experimental results smoothed over discrepancies and, at the same time made it impossible to discern changes in strategies. Therefore, it may be argued that the estimated values of the critical times, T^*, which were constructed graphically, represented an average over several time thresholds, each associated with a given algorithm requiring a certain amount of cognitive workload.

Both the experimental and analytical results confirmed the stability of of the bounded rationality constraint, F_{max}, across similar tasks and across subjects. However, when comparing the experimental and analytical results, it appeared that the stability of $F_{max,3}$ and $F_{max,6}$ over subjects (both distributions are normal) was a more reliable result than the stability of the individual F_{max} across subjects. (The Q^2 value was larger for the F_{max} distribution than for the $F_{max,4}$ and $F_{max,7}$ distributions). This slight difference is derived from the discrepancy between the workload per comparison for trials of three and trials of six tasks. One may conclude however that F_{max} is stable across tasks for each individual, across individuals for each type of task, and finally that F_{max} is stable when considered simultaneously across tasks and across individuals. Considering the nature of the experiment, (i.e., the large size of the input alphabet which did not allow enumeration), the number of different strategies that could be used to perform the task, and the speed at which some of the subjects were capable to perform the task, it can be concluded that the obtained results were very significant.

This experiment is only the first in a series of experiments trying to analyze and quantify the bounded rationality of human decisionmakers under pressure. The task which was analyzed was very basic and included only a single decisionmaker. Research has been undertaken at the Laboratory for Information and Decision Systems at MIT to design multi-person experiments and both validate some of the results obtained in this project on a multiperson level and derive other conclusions on the behavior of the bounded rationality constraint. When considering multi-person organizations, the impact of one DM being overloaded on the performance on the organization as a whole is an important topic to investigate in the context of command and control organizations.

Of course, the key objective of the overall research effort is to design and evaluate organizations carrying out distributed tactical decisionmaking. The results of this project allow now the calibration of the human decisionmaker models for use in the algorithms for the design of organizations that meet performance requirements.

10. REFERENCES

Anderson, N. H. (1974). Information integration theory: A brief survey. In Krantz, Atkinson, Luce, and Suppes (Eds.) *Contemporary developments in mathematical psychology: measurement, psychophysics, and neural information processing*. Freeman and Co.

Anderson, N. H. (1983). *Foundations of information integration*. Academic Press, New York.

Anderson, N. H., and J. C. Shanteau (1970). Information integration in risky decision making. *Journal of Experimental Psychology, 34,* pp.441-451.

Andreadakis, S. K., and A. H. Levis (1987). Accuracy and timeliness in decisionmaking organizations. *Proc. 10th IFAC World Congress*, Pergamon Press, New York.

Bazerman, M. H. (in press). The relevance of Kahneman and Tversky's concept of framing to organization behavior. *Journal of Management.*

Bejjani, G. J. (1985). Information storage and access in decisionmaking organizations. S.M. Thesis, LIDS-TH-1434, MIT, Cambridge, MA.

Ben Zur, H., and S. J. Breznitz (1981). The effects of time pressure on risky choice behavior. *Acta Psychologica, 47,* pp.89-104.

Bettman, J. R., W. J. Johnson and J. W. Payne (1986). Cognitive effort and decision making strategies: A componential analysis of choice. Unpublished manuscript, Fuqua School of Business, Duke University.

Boettcher, K. L., and A. H. Levis (1982). Modeling the interacting decisionmaker with bounded rationality. *IEEE Trans. Systems, Man , and Cybernetics*, SMC-12, pp. 334-344.

Cartwright, D. (1973). Determinants of scientific progress. *American Psychologist, 28,* pp.222-231.

Chyen, G.H.-L., and A. H. Levis (1985). Analysis of preprocessors and decision aids in organizations. *Proc. IFAC/IFIP/IFORS/IEA Conference on Analysis, Design and Evaluation of Man-Machine Systems*, Varese, Italy.

Christensen-Szalanski, J. J. J., and L. R. Beach (1984). The citation bias: fad and fashion in the judgment and decision literature. *American Psychologist, 39,* pp.75-78.

Conant, R. C. (1976). Laws of information which govern systems. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC-6, No. 4.

Cothier, P. (1984). Assessment of timeliness in command and control. S.M. Thesis, LIDS-TH-1391, MIT, Cambridge, MA.

Crozier, W. R. (1978). Evaluating the worth of gambles. *British Journal of Psychology, 69,* pp.179-183.

Donders, F. E. (1983). *Die schnelligkeit psychischer processse.* Archiv Anatomie und Physiologie.

Einhorn, H. J., and R. M. Hogarth (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32,* pp.53-88.

Gopher, D., M. Brickner, and D. Navon (1982). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. *Journal of Experimental Psychology: Human Perception and Performance, 8,* pp.146-157.

Gopher, D., and E. Donchin (1986). Workload: An examination of the concept. In K. Boff, and L. Kaufman (Eds.), *Handbook of Perception and Human Performance.* John Wiley and Sons, New York.

Green, D. M., and J. A. Swets (1966). *Signal detection theory and psychophysics.* John Wiley and Sons, New York.

Greitzer, F. L., and R. L. Hershman (1984). Modeling decision making strategies in simulated AAW operations. *Proc 7th MIT/ONR Workshop on C^3 Systems.* LIDS-R-1437, MIT, Cambridge, MA. pp.145-150.

Greitzer, F. L., S. G. Hutchins and R. T. Kelley (1984). Dual-task performance in a simulated anti-air warfare (AAW) problem. Report NPRDC TR-84-39.

Hall, S. A. (1982). Information theoretic models of storage and memory. S.M. Thesis, LIDS-TH-1232, Laboratory for Information and Decision Systems, MIT, Cambridge, MA.

Hall, S. A., and A. H. Levis (1983). Information theoretic models of memory in human decisionmaking models. *Proc. 6th MIT/ONR Workshop on C^3 Systems*, LIDS-R-1354, MIT, Cambridge, MA.

Hillion, H. P., and A. H. Levis (1987). Timed event-graph and performance evaluation of systems. *Proc. 8th European Workshop on Applications and Theory of Petri Nets,* Zaragoza, Spain.

Hillion, H. P. (1986). Performance evaluation of decisionmaking organizations using timed Petri Nets. S. M. Thesis, LIDS-TH-1590, Laboratory for Information and Decision Systems, MIT, Cambridge, MA.

Hogarth, R. M. (1975). Decision time as a function of task complexity. In Wendt and Vlek (Eds.), *Utility probability, and human decision making.* D. Reidel Publishing Co., Dordrecht, Holland.

Hogarth, R. M. (1980). *Judgment and choice.* John Wiley and Sons, New York.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90,* pp.197-217.

Hogarth, R. M., and S. Makridakis (1981). The value of decision making in a complex environment: An experimental approach. *Management Science, 27,* pp.93-107.

Kahneman, D., P. Slovic, and A. Tversky (Eds.),(1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, England.

Kahneman, D., and A. Tversky (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47,* pp.263-291.

Karkarmar, U. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance, 21,* pp.61-72.

Kelley, R. T., and F. L. Greitzer (1982). Effects of track load on decision performance in simulated command and control operations. Report NPRDC TR-82-21.

Khinchin, A. I. (1957). Mathematical foundations of information theory. Dover, New York.

Kleinmuntz, D. N., and J. B. Thomas (in press). The value of action and inference in dynamic decision making. *Organizational Behavior and Human Decision Processes.*

Levis, A. H. (1984). Information Processing and Decisionmaking Organizations: A Mathematical Description. *Large Scale Systems.* Vol. 7, pp.151-163.

Levis, A. H. (1988). Human organizations as distributed intelligence systems. *Proc. IFAC Symposium on Distributed Intelligence Systems*, Pergamon Press, Oxford.

Lopes, L. L. (1982). Toward a procedural theory of judgment. Technical report, Wisconsin Human Information Processing Program (WHIPP 17), Madison, WI.

Lopes, L. L. (in press). Between hope and fear: The psychology of risk. *Advances in Experimental Social Psychology.*

Lopes, L. L., and J. T. Casey (1987). Tactical and strategic responsiveness in a competitive risk-taking game. Technical report, Wisconsin Human Information Processing Program (WHIPP 28), Madison, WI.

Louvet, A. C., J. T. Casey, and A. H. Levis (1988). Experimental investigation of the bounded rationality constraint. In *The Science of Command and Control,* S. E. Johnson and A. H. Levis, Eds., AFCEA International Press, Washington, DC.

March, J .G.and H. A. Simon (1958). Organizations. John Wiley and Sons, New York.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *Bell J. Economics*, Vol. 9.

Martino, J. P. (1972). *Technological forecasting for decision-making*. New York, NY: Elsevier.

McGill, W. J. (1954). Multivariable information transmission. *Psychometrika*, Vol. 19, No. 2.

Miller, J. G. (1969). Adjusting to overloads of information. *Organizations: Systems, Control, and Adaptation,* Vol. II, J.A. Litterer, Ed., John Wiley and Sons, New York.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive* *psychology*. Freeman, San Francisco, CA.

Newell, A., and H. A. Simon (1972). *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, N.J.

Nisbett, R., and L. Ross (1980). *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall, Englewood Cliffs, N.J.

O'Donnell, R., and F. T. Eggemeier (1986). Workload assessment methodology. In K. Boff, and L. Kaufman (Eds.), *Handbook of Perception and Human Performance*. John Wiley and Sons, New York.

Payne, J. (1982). Contingent decision behavior. *Psychological Bulletin, 92,* pp.382-402.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1986). Adaptive strategy selection in decision making. Unpublished manuscript, Fuqua School of Business, Duke University.

Peterson, J. L. (1981). Petri Net theory and the modeling of systems. Prentice-Hall, Englewood, NJ.

Schneider, S. L., and L. L. Lopes (in press). Reflection in preferences under risk: Who and when may suggest why. *Journal of Experimental Psychology: Human Perception and Performance.*

Schneider, W., S. T. Dumais and R. M. Shiffrin (1983). Automatic and control processing and attention. In R. Parasuraman, R. Davis, and J. Beatty (Eds.), *Varieties of attention*. Academic Press, New York.

Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature, 20,* pp.529-563.

Shannon, C. E., and W. Weaver (1949). The mathematical theory of communication. University of Illinois, Urbana, IL.

Slovic, P. (1962). Convergent validation of risk-taking measures. *Journal of Abnormal and Social Psychology, 65,* pp.68-71.

Slovic, P., B. Fischhoff, and S. Lichtenstein (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* pp.1-39.

Stabile, D. A., and A. H. Levis (1984). The design of information structures: Basic Allocation strategies for organizations. *Large Scale Systems*, Vol. 7.

Stabile, D.A. (1981). The design of information structures: Basic allocation strategies for organizations. S.M. Thesis, LIDS-TH-1098, Laboratory for Information and Decision Systems, MIT, Cambridge, MA.

von Neumann, J., and O. Morgenstern (1947). *The Theory of games and economic behavior.* Princeton NJ: Princeton University. (2nd edition)

Wickens, C. D. (1984). *Engineering psychology and human performance.* Charles F. Merrill, Columbus, OH.

Wierwille, W. W., and J. G. Casali (1983). A validated rating scale for global mental workload measurement applications. Cited in O'Donnell and Eggemeier (1986).

Wright, G. (1985). *Behavioral Decision Making*. Plenum, New York.

Wright, P. L. (1974). The harassed decision maker: Time pressures, distraction, and the use of evidence. *Journal of Applied Psychology, 59,* pp.555-561.

Wright, P. L., and B.Weitz (1977). Time horizon effects on product evaluation strategies. *Journal of Marketing Research, 14,* pp.429-443.

LABORATORY FOR INFORMATION AND DECISION SYSTEMS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

APPENDIX:

EXPERIMENTAL PROGRAM USER GUIDE

(Version 1.1)

by
Jeff T. Casey

July 1987

TABLE OF CONTENTS

I. Introduction

This program is designed to collect from human subjects experimental data concerning the bounded rationality constraint. Technical details of the program can be found in this report or in Louvet et al. (1988). Should modifications of the source code be necessary, an annotated listing is available on the Program Diskette.

II. Hardware and Software Requirements

The program has been run successfully on a Compaq Deskpro Model 2 equipped with an 8087 math co-processor, a monochrome graphics card (640 X 200 pixels), 640K of random access memory, and a monochrome monitor. The program is written in Turbo Pascal version 3.01A. The operating system that has been used is MS-DOS version 2.11. The program has also been run on an IBM PC AT with the 80287 math co-processor and 640K of memory. None of the high resolution graphics capabilities of the AT were used. Therefore, the program should be portable to a wide variety of PC compatible machines.

III. Setting Up

Two diskettes are required: the **program diskette** and a **data diskette**. The following Turbo Graphics Toolbox files must be present on the program diskette:

TYPEDEF.SYS,
GRAPHIX.IBM,
KERNEL.SYS,
WINDOWS.SYS, and
4X6.FON.

In order to prevent loss of data, the data diskette must have enough free space to store the data file. For the experiment reported in Louvet et al. (1987), approximately 13K bytes were needed per file (7K for practice session files), where one file contains the data from a single experimental session. *Data files should be backed-up immediately following each session.*

The procedure for setting up to run the program differs on machines having two floppy disk drives versus those having a single floppy and a hard disk. On machines with two floppy disk drives, insert the program diskette in the primary drive (Drive a) and a data diskette in the remaining drive (Drive b). On machines with a hard disk, first insert the program diskette and copy its contents onto the hard disk. Then replace the program diskette with a data diskette. If the experimental program is already resident on the hard disk, the copy step may be omitted and only a data diskette is needed. On either system, execute the compiled program file from MS-DOS by typing EXPER <return>.

IV. Sample Session

The following is a sample dialogue between the program and the experimenter. The information generated by the program is shown in *italics*. The permissible responses are shown in brackets.

Use E to exit program at input statements.
Is this going to be a demonstration only? [Y,N,E]

This option is used for the initial demonstration of the program to subjects. In addtition, subjects should be given a few minutes of practice in the demonstration mode prior to the start of each session. As described below, the demonstration mode permits the program to be aborted at any time. Entering 'E' in response to any query prior to the start of the session will cause the program to abort.

Use default parameters? [Y,N,E]

The parameters referred to here are the minimum time per comparison, maximum time per comparision, increment size for time per comparison, the various numbers of threats to be used, the number of iterations between breaks (rest periods), and the number of repetitions of the entire iterate/break cycle. The parameters are different for demonstration, practice and actual experiment. (The time per comparison values are greatest for demonstration and least for the actual experiment.) Refer to the source code if it is necessary to change permanently the default parameters. (The name of the releveant procedure is GETINFO.)

Display Info? [Y,N,E]

If this question is answered 'N', the information underlined below will not appear.

Otherwise this information will appear, whether or not the default parameters are used. The purpose of this option is to prevent subjects from seeing, and perhaps misinterpreting, information concerning the total number of comparisons, approximate duration of session, etc. This option does not affect the subsequent queries.

If the "use default parameters" query is answered 'N', the experimenter will be queried concerning each of the primary experimental parameters listed above:

time values are REAL, but must be in multiples of 0.05 seconds.
minimum number of seconds per comparison: [real number]
maximum number of seconds per comparison: [real number]
change time per comparison in what size steps (in seconds)? [real number]

Note that the difference between the minimum and maximum time per comparison must be an even multiple of the step size.

number of different time per comparison values = X
how many different numbers of targets (INTEGER)? [integer, e.g. 2]
total number of trials = X
(for one ascending & one descending sequence)
enter number of targets [1]: [integer < 13]
enter number of targets [2]: [integer < 13]

Normally, one complete revolution of the clock consumes 30 seconds. If the maximum time per trial (i.e., the product of the maximum number of seconds per comparison and the maximum number of comparisons--the maximum number of targets minus one) is greater than thirty seconds, a minor modification must be made to the source code. The constant THIRTYSECCLK must be set to false. This change will cause one clock revolution to consume 60 seconds.

total number of comparisons per replication = X
approximate total duration of one replication = X minutes
replicate the series of trials specified above how many times (INT)? [integer]

This is the number of replications before a break. (A replication consists of o. e ascending and one descending iteration for each number of targets.) It is desirable that this value be adjusted to allow a break at least every 10 minutes.

replicate the resulting series of trials how many times (INT)? [integer]

This is the number of repetitions of the entire iterate/break cycle (i.e., the number of breaks in the course of the session). A break will not be inserted at the end of the session.

<u>*total number of trials = X*</u>
<u>*total number of comparisons = X*</u>
<u>*approximate total duration of experiment = X minutes*</u>

If this is not a demonstration session, then data will be stored on disk. In this case, the experimenter is queried for a subject number and a session number:

subject number: [integer]
session number: [integer or P]

The 'P' option is used to indicate that this is a practice session. Practice data <u>are</u> stored on disk. Actual experimental sessions are numbered sequentially.

If the machine in use has a hard disk and one floppy, and this is not a demonstration session, the following message will appear:

Insert diskette for drive B: and strike
any key when ready

In this case, simply press return. However, if this message reappears, or if a two-floppy disk system is in use, check to see that the data diskette is in place.

If a data file already exists on the diskette for the subject and session numbers specified, the following message will appear:

Data file for subject A session B already exists. OVERWRITE? [Y,N,E]

'Y' will cause the existing file to be overwritten. 'N' will cause the program to abort. Note that there are many ways of losing data despite this safeguard. Thus it is important to keep an up-to-date log of subjects and sessions.

After a pause the following query will appear:

run experiment? [Y,N,E]

The subject should be seated and poised at the keyboard before a response is made to the above query. 'Y' initiates the session. 'N' causes the program to abort. For practice and real sessions, this is the last opportunity to abort the program without re-booting. Demonstration sessions may be aborted at virtually any time by pressing the space bar, followed by 'E'. Normal termination of a session is indicated by a "thank you for your help" message. The data file is closed prior to this message. Pressing the space bar at this point will cause the program to terminate.

V. Instructions to Subject

Instructions similar to the following (used by Louvet et al., 1988) should be provided to the subject in written form:

> In this experiment, we are attempting to measure how much information people are able to process accurately in a fixed amount of time.
>
> The experiment involves a computer game. The way the game works is this: The large circle represents a radar screen. You are located at the dot in the center of the radar screen. There are several "targets" or "threats" (e.g., enemy aircraft) converging simultaneously on your location. Your task is to determine which threat will reach you first (e.g., so that it can be intercepted). The threats will be shown on the radar screen two at a time. For each threat, two pieces of information will be provided--the distance and the speed. This information will be presented as a fraction for each threat--the numerator is the distance and the denominator is the speed. Thus the threat which has the *smallest ratio* is the one that will reach you first. You will enter your responses into the computer by using the arrow keys on the numeric keypad on the right side of the keyboard. The experimenter will show you how to use these keys.
>
> Once you have selected the threat with the smaller ratio, the *other* threat will disappear and a new one will appear which has a different ratio. Once again you choose the threat having the smaller ratio. All of the threats that are converging upon you are represented in the box to the left of the radar screen. Each time a

new threat appears on the radar screen, one will disappear from this box. Thus the number of threats you have left to compare is shown by the number of DI/SP's that remain in the box. After you have compared the final two threats and entered your response, the correct answer (the threat with the smallest ratio) will flash. If you were correct, then the one target remaining on the screen will be the one that flashes. Then a new *trial* will begin; that is, a new set of threats will appear in the box and you will repeat the process.

The amount of time you have to compare the ratios of all of the threats will change systematically from trial to trial. The clock shows how much time you have and how much of the time has elapsed thus far. One revolution of the clock takes *30 seconds* (not 1 minute). The clock always starts at 12 o'clock. If, for example, the other hand is at 6 o'clock, this indicates that you have 15 seconds to compare all of the threats. The moving hand shows elapsed time. If you have not yet finished when time expires, the computer will go on to the next trial.

At the beginning of each trial the clock's second hand will flash to indicate the amount of time allotted for the trial. Be sure to notice how much time is indicated. With practice you will be able to use this information to pace yourself and take full advantage of the amount of time available. The amount of time allotted for each trial will be relatively large at first and will then gradually *decrease* to a minimum value. When the minimum value is reached, the time per trial will begin *increasing* and continue increasing until a maximum value is reached.

As time pressure increases, you will sometimes have too little time to compare the ratios carefully. Unless you are confident that your response will be correct, it is better to risk letting time run out before you finish all of the comparisons. You will hear a low-pitched tone whenever you make an *incorrect* response.

Every so often, the number of threats in the box will change and, therefore, the number of comparisons you have to make will change. The entire box will flash to indicate that the number of threats is about to change. When this happens, you will be allotted proportionately more or less time, as indicated by the clock.

A two-second, high-pitched tone indicates that it is time for a break. The program will pause until you press the space bar.

The experimenter will be happy to answer questions at any time.

While the subject reads the instructions, a "frozen" in-progress trial should be present on the screen. This can be accomplished by running the program in the demonstration mode. After the first trial has begun, press the space bar to freeze the program. Press the space bar again to make the program resume. When the subject has finished reading the instructions, the experimenter should reiterate the major points and help the subject through the first two or three trials.

VI. Data File Format

Data file names have the following format:

xSyNz.D ,

where x is a one-digit integer indicating the number of the experiment (e.g., the experiment reported by Louvet et al. was experiment 1), y is a two-digit integer indicating the subject number, and z is a one-digit integer indicating the session number or the letter 'P' (indicating a practice session). The data files are text files (e.g., they can be edited using the Turbo Pascal Text Editor). Each data file line contains the data for one trial. The format of each line is shown in Table 1.

Table 1: Format of each data file line

Variable	Type	Field Specification
Time per comparison	real	1.2*
Number of targets	integer	2
Direction of sweep	integer	1
Performance	integer	1
blank space	--	1
Distance of target 1	integer	2
Speed of target 1	"	"
.	"	"
.	"	"
Distance of target n**	"	"
Speed of target n	"	"
blank space	--	1
Reponse to comparison 1	integer	1
.	"	"
.	"	"
Response to comparison n-1	"	"
blank space	--	1
Response time to comparison 1	integer	4
.	"	"
.	"	"
Response time to comparison n-1	"	"

* Format is a.b, where a is number of digits to left of decimal and b is number of digits to right of decimal.

**n = number of targets.

Time per comparison is in seconds. A value of 1 for direction of sweep denotes a series descending in terms of time per comparison, while a value of 2 denotes an ascending series. (A descending series always preceeds an ascending series.) A value of 0 for performance indicates an incorrect reponse, a value of 1 indicates a correct response, and a value of 2 indicates that the comparisons were not completed before time ran out.

The distance and speed values are in the order in which they were presented to the subject. Thus the values for the targets presented for comparison 1 (i.e., targets 1 and 2) are encountered first as the information is read from left to right. Each response to comparison value is the number of the target that was selected on that comparison. Thus the first response to comparison value will always be 1 or 2, the second 1, 2 or 3, etc. A response to comparison value of 0 indicates that time ran out before a response was made to that comparison. The correctness of any comparison can be determined by finding the actual correct response for the comparison from the distance and speed information and comparing the result with the corresponding response to comparison value. The response time to comparison values are in milliseconds. The time is not cumulative; it is measured from the time the new ratio(s) for the comparison appear on the screen.

VII. Reference

Louvet, A. C., J. T. Casey and A. H. Levis (1988). Experimental investigation of the bounded rationality constraint. *Proc. 1987 Symposium on C3 Research,* National Defense University, Ft. McNair, Washington, DC.